



Departamento de Matemática

MESTRADO EM MODELAÇÃO ESTATÍSTICA E ANÁLISE DE DADOS

Uma Caracterização dos Agregados Familiares

Portugueses

Estudo Comparativo com Espanha

Dissertação de Mestrado sob a orientação do Professor Doutor Paulo Infante

Departamento de Matemática – Universidade de Évora

Patrícia M. S. Ribeiro | Évora/2010



Departamento de Matemática

MESTRADO EM MODELAÇÃO ESTATÍSTICA E ANÁLISE DE DADOS

Uma Caracterização dos Agregados Familiares

Portugueses

Estudo Comparativo com Espanha

Dissertação de Mestrado sob a orientação do Professor Doutor Paulo Infante

Departamento de Matemática – Universidade de Évora

Patrícia M. S. Ribeiro | Évora/2010

Melhorias no crescimento do país não devem ser traduzidas no consumo das famílias, pois continuam com índices de endividamento e consolidação muito elevados, tendo alcançado novos máximos históricos.

Relatório do FMI, Outubro de 2005

Agradecimentos

Não posso deixar de agradecer aos principais responsáveis por ter chegado até aqui que, são os meus filhos...pelo respeito, que demonstraram no decorrer deste mestrado que inúmeras vezes causou a ausência da mãe e do pai, pela partilha, que nos levou a “estudar” em conjunto...mesmo quando esta palavra, ainda apenas, significava para eles, ver as imagens de um livro...ou fazer um simples desenho, pelas alegrias que me proporcionaram e que tantas vezes serviram de empurrão para prosseguir, e por todos os momentos.

Chegar até aqui, fez-me espreitar para o dia em que decidi que queria participar neste desafio e comprovei que as pessoas que mais me apoiaram, continuam ao meu lado. Todo o meu agradecimento é para elas.

Índice

Índice de tabelas	vi
Índice de figuras	viii
Resumo	x
Abstract	xi
Capítulo I	1
1 Introdução	1
1.1 Metodologia do Inquérito	2
1.1.1 Caracterização metodológica do inquérito	2
1.1.2 Desenho dos questionários	3
1.1.3 Classificação das despesas de consumo	3
1.1.4 Anualização dos dados	4
1.1.5 Amostragem	5
1.1.6 Recolha de dados	6
1.1.6.1 Recolha por auto-preenchimento	6
1.1.6.2 Recolha por entrevista	6
1.1.7 Caracterização das Variáveis	6
1.1.7.1 Despesas	6
1.1.7.2 Rendimentos	7
1.1.7.3 Outras variáveis em estudo	7
1.2 Metodologia estatística	8
1.2.1 Análise exploratória de dados	8

1.2.2 Curva de Lorenz	9
1.2.3 Índice de GINI	9
1.2.4 Modelos Lineares Generalizados	10
1.2.4.1 Regressão logística	11
1.2.4.1.1 Características do modelo	11
1.2.4.1.2 Interpretação dos coeficientes	14
1.2.4.1.3 Avaliação da bondade do ajustamento do modelo	15
1.2.4.1.4 Selecção de variáveis	17
1.2.4.1.5 Diagnóstico do Modelo	19
Capítulo II	23
2.1 Análise dos dados	23
2.1.1 Estatística Descritiva	23
2.1.2 Estudo Gráfico da Distribuição das Variáveis	29
2.1.3 Classes de variáveis	30
2.1.4 Curva de Lorenz, Índice de GINI	33
2.1.5 Análise Descritiva de Outras Variáveis	42
Capítulo III	47
3. Modelação Estatística da Despesa dos Agregados	47
3.1 Modelo de Regressão Logística para Portugal	47
3.1.1. Bondade do Ajustamento e Diagnóstico	52
3.1.2. Interpretação dos Coeficientes	57

3.2 Modelo de Regressão Logística para Espanha	59
3.2.1 Bondade do Ajustamento e Diagnóstico	60
3.2.2 Interpretação dos Coeficientes	63
3.3 Modelo de Regressão Logística para Portugal e Espanha	64
3.3.1 Bondade do Ajustamento e Diagnóstico	66
3.3.2 Interpretação dos Coeficientes	70
Capítulo IV	71
4. Considerações finais	71
5. Bibliografia	74
Anexo I – Questionário utilizado no Inquérito às Despesas das Famílias (2005/2006)	76
Anexo II – Modelo de Portugal (scripts)	106
Anexo III – Modelo de Espanha (scripts)	114
Anexo IV – Modelo de Portugal e Espanha (scripts)	133

Índice de Tabelas

Tabela 1: Alguns modelos lineares generalizados	11
Tabela 2: Poder discriminativo da área sob a curva de ROC	22

Tabela 3: Medidas amostrais relativas às variáveis sobre despesa, em Portugal e Espanha	24
Tabela 4: Medidas amostrais relativas às variáveis sobre rendimentos, em Portugal e Espanha	27
Tabela 5: Tabela de codificação das variáveis referentes aos dados de Portugal e Espanha	31
Tabela 6: Classes da variável rendimento total, valores em € (Portugal)	33
Tabela 7: Distribuição do rendimento total, valores apresentados em € (Portugal)	34
Tabela 8: Distribuição do rendimento total, valores das classes apresentados em € (Espanha)	34
Tabela 9: Distribuição do rendimento monetário, valores apresentados em € (Portugal)	36
Tabela 10: Distribuição do rendimento não monetário, valores apresentados em € (Portugal)	37
Tabela 11: Distribuição da variável despesa total, valores apresentados em € (Portugal)	37
Tabela 12: Distribuição da variável despesa total, valores apresentados em € (Espanha)	37
Tabela 13: Distribuição da variável despesa monetária, valores apresentados em € (Portugal)	39
Tabela 14: Classes da variável despesa monetária, valores das classes apresentados em € (Espanha)	39
Tabela 15: Distribuição da variável despesa não monetária, valores apresentados em € (Espanha)	40
Tabela 16: Distribuição da variável despesa não monetária, valores apresentados em € (Espanha)	41
Tabela 17: Comparação dos resultados do Índice de Gini obtidos através dos valores reais e em classes, para Portugal e Espanha	42
Tabela 18: Comparação dos resultados do valor da média e do desvio padrão obtidos através dos valores não agrupados e agrupados em classes para Portugal	42
Tabela 19: Tabela de frequências para o número de pessoas por agregado familiar	44
Tabela 20: Resultados do ajustamento de modelos com cada variável Individualmente	47
Tabela 21: Modelos de regressão logística ajustados para Portugal e respectivas tabelas de resultados	49
Tabela 22: Modelo de regressão logística ajustado para Portugal	51

Tabela 23: MRL final, para Portugal	52
Tabela 24: Valores de sensibilidade e especificidade obtidos, com diferentes pontos de corte para o modelo de Portugal	56
Tabela 25: Resultados do ajustamento de modelos com cada variável individualmente (Espanha)	59
Tabela 26: MRL final, para Espanha	60
Tabela 27: Valores de sensibilidade e especificidade obtidos, com diferentes pontos de corte para o modelo de Portugal	63
Tabela 28: Resultados do ajustamento de modelos com cada variável individualmente (Portugal+Espanha)	65
Tabela 29: MRL final, para Portugal e Espanha	65

Índice de Figuras

Figura 1: Exemplo de curva ROC	22
Figura 2: Histogramas das variáveis despesa total, despesa monetária e despesa não monetária, em Portugal e Espanha	25
Figura 3: Caixas de bigodes das variáveis despesa total, despesa monetária e despesa não monetária, em Portugal e Espanha	26

Figura 4: Histogramas das variáveis rendimento total, em Portugal e Espanha e despesa monetária e despesa não monetária, em Portugal	28
Figura 5: Caixas de bigodes das variáveis rendimento total, em Portugal e Espanha e despesa monetária e despesa não monetária, em Portugal	29
Figura 6: Q-Q plot das distribuições Normal, Lognormal e Gama, variável despesa total (Portugal)	29
Figura 7: Curva de Lorenz, com os valores da variável Rendimento total, em Portugal e Espanha	35
Figura 8: Comparação entre razões da mediana com os percentis 5%, 10%, 25%, 75%, 90% e 95% para o rendimento total em Portugal e Espanha	36
Figura 9: Comparação entre razões da mediana com os percentis 5%, 10%, 25%, 75%, 90% e 95% para a despesa total em Portugal e Espanha	38
Figura 10: Comparação entre razões da mediana com os percentis 5%, 10%, 25%, 75%, 90% e 95% para a despesa monetária em Portugal e Espanha	40
Figura 11: Comparação entre razões da mediana com os percentis 5%, 10%, 25%, 75%, 90% e 95% para a despesa não monetária em Portugal e Espanha	41
Figura 12: Distribuição da variável grau de urbanização por categorias	43
Figura 13: Distribuição da variável alojamento por categorias	44
Figura 14: Histograma para o índice de conforto total	45
Figura 15: Representação dos coeficientes ajustados pelo modelo para as categorias da covariável continua conftotal categorizada	52
Figura 16: Resíduos de Pearson	53
Figura 17: Resíduos <i>Deviance</i>	54
Figura 18: Resíduos DfBetas	55
Figura 19: Distância de Cook	56
Figura 20: Curva ROC para o modelo logístico das despesas dos agregados portugueses	57
Figura 21: Resíduos <i>Deviance</i> para o modelo de Espanha	61

Figura 22: Resíduos Dfbetas para o modelo de Espanha	62
Figura 23: Distância de Cook para o Modelo de Espanha	62
Figura 24: Curva ROC para o modelo logístico das despesas dos agregados espanhóis	63
Figura 25: Resíduos <i>Deviance</i> para o modelo conjunto	67
Figura 26: Resíduos Dfbetas para o modelo conjunto	68
Figura 27: Distância de Cook para o Modelo conjunto	69
Figura 28: Curva ROC para o modelo logístico das despesas dos agregados portugueses e espanhóis	70

Resumo: Uma caracterização dos agregados familiares portugueses – Estudo Comparativo com Espanha

No contexto actual assume grande importância conhecer e analisar estatisticamente determinadas características sobre os agregados familiares. Considerando os dados existentes no Instituto Nacional de Estatística em Portugal, obtidos através do inquérito às despesas das famílias (IDEF 2005/2006) e os dados existentes no Instituto Nacional de Estatística em Espanha, nesta dissertação caracterizam-se os agregados familiares de ambos os países relativamente à despesa anual, rendimento anual, número de pessoas por alojamento, regime de ocupação e alguns indicadores de conforto. Analisa-se também a concentração dos rendimentos e das despesas das famílias em ambos os países, sendo de realçar uma grande proximidade dos valores obtidos.

Através de modelos de regressão logística modelou-se a despesa dos agregados familiares portugueses e espanhóis em função do rendimento anual, do número de pessoas por alojamento, do regime de ocupação e de um índice de conforto. Para um perfil fixo relativamente a estas variáveis, podemos concluir que os agregados espanhóis têm 10% menos de possibilidades de terem uma despesa acima da mediana que os agregados portugueses.

Palavras-chave: despesa, agregados familiares, modelação de dados

Abstract: A characterization of the Portuguese's households – Comparative study with Spain

Nowadays is very important to know and analyse statistically some households' characteristics. According to the data sets available in Statistics Portugal, obtained from Inquérito às Despesas das Famílias (IDEF 2005/2006) and in Statistics Spain, in this dissertation, households in both countries are characterized about annual expenditure and income, number of persons by household, occupation scheme and some well being indicators. Concentration of households' income and expenditure are also analyzed and the proximately obtained in both countries should be emphasized.

Logistic regression models were used to model the expenditure of households living in Portugal and in Spain, using annual income, number of persons in the household, occupation scheme and a well being index as covariates.

Having a fixed profile related to these variables, can be conclude that the Spanish households have 10% less possibilities of having a expenditure above the median when compared to the Portuguese households.

Key-words: expenditure, income household, modelling data sets.

Capítulo I

1 Introdução

Portugal tem vindo a reunir esforços para melhorar as políticas em matéria de coesão social através da tentativa de redução da pobreza infantil e promoção para inclusão activa das pessoas mais desfavorecidas na sociedade. Por outro lado, assiste-se ao aumento da frequência de situações sociodemográficas atípicas que contemplam, entre outros, alteração da constituição dos agregados familiares ou tendência crescente do desemprego. Neste contexto, pensamos que assume uma grande importância conhecer e analisar estatisticamente determinadas características sobre as famílias residentes em Portugal comparando com as de países de estados membros da UE e, em particular, com a vizinha Espanha.

Considerando os dados existentes no Instituto Nacional de Estatística (Portugal), obtidos através do inquérito às despesas das famílias (IDEF 2005/2006) e os dados existentes no Instituto Nacional de Estatística (Espanha), nesta tese realiza-se um estudo que procura caracterizar e comparar os agregados familiares portugueses e espanhóis, considerando variáveis como a repartição dos agregados, despesa anual média, rendimento anual médio, número de pessoas por alojamento, regime de ocupação e alguns indicadores de conforto.

A publicação do INE (2008), que constitui o relatório sobre o IDEF 2005/06 apresenta um vasto conjunto de resultados estatísticos, apurando indicadores sobre a distribuição do rendimento e estrutura das despesas dos agregados familiares e sobre as respectivas condições de conforto. A publicação indicada foi a única onde se conseguiu obter informações acerca da abordagem que pretendíamos utilizar, ao nível nacional. Foram pesquisadas informações sobre a situação existente em Portugal durante o período de análise e, nos relatórios de organismos internacionais, nomeadamente o FMI (Outubro, 2005), já eram apresentados indícios que os agregados familiares apresentavam sinais de endividamento.

Após uma análise exploratória das variáveis em estudo, procedemos à comparação entre os dois países, ao nível da concentração, dos rendimentos e das despesas. No próximo capítulo, para além da descrição da metodologia apresentamos os aspectos teóricos principais das ferramentas estatísticas usadas no trabalho e a análise efectuada aos dados.

O terceiro capítulo é dedicado à análise univariada, multivariada, verificação de pressupostos de ajustamento e respectivas interpretações do modelo de regressão logística para os dados relativos a Portugal. O mesmo procedimento foi efectuado para Espanha.

No quarto capítulo serão apresentadas as considerações finais.

1.1 Metodologia do Inquérito

1.1.1 Caracterização metodológica do inquérito

Os resultados provenientes o IDEF inserem-se numa série de dados estatísticos sobre orçamentos familiares, iniciada na década de '60, do séc. XX e realizada com periodicidade quinquenal.

As estatísticas sobre orçamentos familiares constituem informação conduzida a nível comunitário – *household budget survey* – sobretudo no que se refere à proposta de metas de harmonização para o período de recolha e para a classificação do consumo por objectivo, que serve de referência para o registo das despesas.

A informação relativa a Portugal reflecte os dados recolhidos entre Outubro de 2005 e Outubro de 2006, junto de 10403 famílias que colaboraram no registo diário de todas as despesas efectuadas no decorrer da quinzena para a qual foram seleccionadas.

Os dados provenientes do mesmo inquérito, em Espanha, resultam da colaboração de 19435 famílias.

O IDEF é um inquérito associado a um questionário que inclui cadernetas para o preenchimento, pelas famílias seleccionadas, de todas as despesas familiares e individuais durante duas semanas.

O questionário pode ser consultado no Anexo I deste trabalho.

De entre os principais objectivos deste inquérito, nos países da União Europeia, podemos distinguir os seguintes:

- Obtenção de estimativas das despesas de consumo global das famílias, bem como a sua classificação de acordo com diversas variáveis do agregado;
- Obtenção de estimativas de alterações interanuais do consumo total do agregado;
- Estimativa do consumo em quantidades físicas, de determinados bens alimentares, para o conjunto nacional.

Além dos objectivos prioritários, podemos ainda destacar outros relacionados com as necessidades concretas do inquérito, como:

- Estimativa da despesa como instrumento para obtenção de valores do consumo privado, na contabilidade nacional;
- Estimativa da estrutura de ponderadores, a partir da despesa, necessários para o cálculo do índice de preços no consumidor (IPC).

1.1.2 Desenho dos questionários

Os questionários, de ambos os países, são compostos por quatro módulos:

O Módulo I procede à caracterização do alojamento, do agregado doméstico privado e do(s) indivíduo(s) – a informação foi recolhida através de entrevista directa com computador (CAPI: *computer assisted telephone interview*) e reporta ao momento da entrevista, com excepção nos rendimentos em que foi considerado o ano fiscal anterior àquele em que ocorreu a quinzena de entrevista;

Os módulos II e III correspondem ao diário de consumo do agregado (II) e ao diário de consumo do indivíduo (III) – os dados foram obtidos por auto-preenchimento ao longo da quinzena de entrevista;

O módulo IV refere-se à recolha retrospectiva dos consumos geralmente realizados com periodicidade mensal, trimestral ou anual e recebimentos gratuitos e a títulos de salário, apelando-se respectivamente para a memória dos consumos efectuados durante os 30 dias anteriores à quinzena da entrevista, durante os três meses anteriores à quinzena da entrevista e no decurso dos doze meses anteriores à quinzena de entrevista – dados recolhidos através de entrevista directa sem computador (PAPI: *personnel and pencil assisted interview*).

1.1.3 Classificação das despesas de consumo

A descrição detalhada associada a cada despesa de consumo, registada nos diários, foi recolhida e analisada com vista à codificação no quadro da classificação do consumo individual por objectivo (COICOP acrónimo inglês: *Classification of individual consumption by purpose*).

As despesas de consumo obtidas nestes inquéritos não se referem apenas ao fluxo monetário que cada família destina e o que cada um dos seus membros paga por determinado bem ou serviço, considerados como consumo final, mas pretende incluir também o valor do consumo efectuado pelas famílias em regime de auto-consumo, auto-abastecimento, salário em espécie ou alimentação gratuita.

O registo das despesas de consumo de um ano inteiro, com o detalhe necessário, só é possível para os bens e serviços adquiridos com menor frequência, visto que o efeito de memória é considerado penalizador para os consumos repetidos. Assim, com base no conhecimento acumulado em experiências anteriores, foram classificados todos os bens e serviços que constituem a COICOP conforme, por um lado, o período de tempo sazonal possível de ser lembrado e, por outro lado, o período de tempo razoável possível de ser lembrado e, finalmente, a frequência com que habitualmente os agregados procedem à sua aquisição.

Utilizaram quatro tipos de periodicidade para as despesas de consumo:

Anual | aplicável a bens ou serviços geralmente adquiridos com frequência reduzida, em que seria expectável uma resposta correcta para os últimos doze meses imediatamente anteriores à entrevista, como por exemplo, despesas com serviços de saneamento, aquisição de electrodomésticos, serviços hospitalares, aquisição de veículos ou seguros;

Trimestral | aplicável aos bens ou serviços adquiridos várias vezes no ano, mas sem periodicidade mensal, como é o caso das despesas com vestuário, calçado, reparação e conservação da habitação, utensílios domésticos, transportes aéreos ou jogos e brinquedos;

Mensal | aplicável nas despesas efectuadas mensalmente, vulgarmente de natureza fixa, como acontece com as despesas relativas a arrendamentos, abastecimentos de água, electricidade ou gás.

Quinzenal | aplicável às despesas com bens e serviços adquiridos com frequência, como por exemplo, alimentação, bebidas, tabaco, artigos domésticos não duradouros, combustíveis ou despesas em restaurantes e cafés.

A informação relativa aos bens e serviços agrupados nos tipos anual, trimestral, e mensal foi obtida por recolha retrospectiva. No caso do tipo quinzenal recorreram ao registo diário ao longo da quinzena de observação.

1.1.4 Anualização dos dados

Os dados sobre despesas de bens ou serviços cuja classificação, de acordo com a COICOP, ficou associada a periodicidade infra-anual (quinzenal, mensal ou trimestral) foram anualizados através da aplicação de um factor multiplicativo (FM) que tem em conta o número de períodos no ano: 26 no caso da periodicidade ser quinzenal, 12 no

caso da periodicidade mensal e 4 no caso de consumos a que está associada periodicidade trimestral e cuja formula é a seguinte:

$$FM=T/t$$

sendo T a duração do período de estudo (365 dias) e t o respectivo período de referência, ambos medidos em número de dias.

Para as despesas do último recibo FM é dado pelo número de vezes que o recibo foi pago, nos últimos 12 meses.

Os dados sobre rendimento foram recolhidos por referência ao ano fiscal anterior àquele em que ocorreu a quinzena da entrevista. Os dados sobre rendimento recolhidos em 2005, como tal reportados a 2004, foram actualizados para 2005 através da aplicação da taxa média de crescimento do índice de preços no consumidor (IPC) nos últimos doze meses, obtida em Dezembro de 2005.

1.1.5 Amostragem

As amostras destes inquéritos correspondem à população residente em território nacional (Continente e Regiões Autónomas, válido para ambos os países em estudo), tanto considerada individualmente, como organizada em agregados domésticos privados, tendo sido excluídos os indivíduos residentes em alojamento colectivos.

A recolha de dados sobre orçamentos familiares desenvolve-se normalmente durante um ano completo. De modo a minimizar os efeitos sazonais nos resultados do inquérito, há a preocupação de assegurar uma razoável dispersão temporal e geográfica das unidades de alojamento da amostra. Deste modo, considerando que o período de observação de cada agregado familiar era de duas semanas, as unidades de alojamento foram distribuídas de forma mais ou menos uniforme por 26 períodos idênticos (quinzena).

Em Portugal, o período de recolha de dados foi de um ano, sendo cada agregado objecto de observação directa durante uma quinzena. A recolha de informação junto das famílias decorreu de 10 de Outubro de 2005 a 8 de Outubro de 2006.

Não dispomos deste tipo de informação para Espanha.

1.1.6 Recolha de dados

1.1.6.1 Recolha por auto-preenchimento

Foram entregues diários de dois tipos para auto-preenchimento por parte dos membros do agregado doméstico privado com 15 ou mais anos (ou 10 anos, com autorização dos detentores do poder paternal).

O diário de consumo do agregado (Módulo II) foi entregue ao indivíduo que melhor conhece as compras do agregado em geral, inscrevendo diariamente todas as despesas com a aquisição de bens e serviços, quer se destinem ao consumo do agregado, quer sejam para oferta a outros agregados ou instituições, os recebimentos gratuitos, o auto-consumo e o auto-abastecimento ocorridos durante o período de observação (uma quinzena).

O diário de consumo do indivíduo (módulo III) foi entregue a cada membro do agregado com idade igual ou superior a 15 anos, inscrevendo diariamente todas as despesas de carácter pessoal ocorridas durante o período de observação (compras, recebimentos gratuitos e auto-abastecimento).

1.1.6.2 Recolha por entrevista

A recolha por entrevista passou por várias etapas:

- Entrevista directa com computador ao representante do agregado, para a recolha de dados sobre o alojamento, o agregado doméstico privado e as condições de conforto (Módulo I);
- Entrevista directa com computador a todos os membros do agregado para a recolha de dados demográfico e sobre trabalho e rendimento (Módulo I), sendo que no caso dos menores de 15 anos foi obtida por procuração (proxy);
- Entrevista para preenchimento nos diários de recolha retrospectiva (mensal, trimestral e anual) das despesas com compras de bens e serviços, auto-abastecimento, recebimentos gratuitos e a título de salário (Módulo IV).

1.1.7 Caracterização das variáveis

1.1.7.1 Despesas

A **Despesa Total** (DespTot) é constituída pelo somatório da Despesa Monetária (DespMonet) e da Despesa Não Monetária (DespNaoMonet).

Despesa Monetária agrega todas as compras de bens e serviços, efectuadas no país ou no estrangeiro, que são para consumo imediato do agregado, oferta ou armazenamento, abrangendo um período de referência retroactivo até aos 12 meses anteriores à quinzena da entrevista.

As aquisições são consideradas pelo seu valor total, independentemente do modo ou momento do pagamento.

Despesa Não Monetária inclui o auto-consumo (bens alimentares e outros de produção própria), o auto-abastecimento (bens ou serviços obtidos, sem recurso a pagamento, de estabelecimento explorado pelo agregado), a auto-locação (auto-avaliação pelos agregados proprietários ou usufrutuários de alojamento gratuito de valor hipotético de renda de casa), recebimentos em géneros e salários em espécie.

1.1.7.2 Rendimentos

O **Rendimento Total** (RendTot) é constituído pelo somatório do **Rendimento Monetário** (RendMonet) e do **Rendimento Não Monetário** (RendNaoMonet).

Rendimento Monetário Líquido agrega os rendimentos obtidos por cada membro agregado, provenientes do trabalho (por conta própria e/ou de outrem), de propriedade e capital, de pensões (nacionais ou procedentes do estrangeiro), de outras transferências sociais (como por exemplo, apoio à família, à habitação, ao desemprego, doença, invalidez, educação, inclusão social, entre outros) e de outras transferências privadas (de agregados domésticos privados ou outras transferências não especificadas) aos quais foram deduzidos impostos.

Rendimento Não Monetário, analogamente à despesa não monetária, abrange o auto-consumo, o auto-abastecimento, a auto-locação, recebimentos em géneros e salários em espécie.

1.1.7.3 Outras variáveis em estudo

Tipologia do agregado familiar (npagr) é constituída pelo conjunto de pessoas que residem no mesmo alojamento e cujas despesas são suportadas em conjunto, independentemente da existência ou não de grau de parentesco.

Desagregação territorial através de tipologia de áreas urbanas (gurb): Área predominantemente rural (APR), área predominantemente urbana (APU), área mediamente urbana (AMU).

Grau de ocupação dos alojamentos (regocup): Proprietário, sem crédito à habitação; proprietário, com crédito à habitação, arrendatário (ou subarrendatário), com renda a

preços de mercado; arrendatário (ou subarrendatário), com renda inferior ao preço de mercado, alojamento cedido gratuitamente ou a título de salário.

Indicadores de conforto do alojamento (conftotal, confbasico), apenas disponíveis para Portugal.

- | | | |
|------------------------------------|--|---|
| 1. Água canalizada | 17. Gravador de cassetes de áudio | 28. Outro aparelho de aquecimento de água |
| 2. Electricidade | 18. Rádio | 29. Desumidificador eléctrico |
| 3. Gás canalizado | 19. Gira-Discos | 30. Telefone – rede fixa |
| 4. Sistema de esgotos | 20. Vídeo | 31. Telefone – rede móvel |
| 5. Inst. sanitária completa | 21. Câmara de vídeo | 32. Computador pessoal, de secretária ou portátil, com ligação à internet |
| 6. Forno ou placa | 22. Equipamento fotográfico | 33. Computador pessoal, de secretária ou portátil, sem ligação à internet |
| 7. Frigorífico | 23. Consolas de jogos (com ou sem leitor de DVD) | 34. Televisão por cabo ou por satélite |
| 8. Microondas | 24. Máquina de costura | 35. Antena parabólica |
| 9. Arca congeladora | 25. Aparelho de ar condicionado | 36. Aparelho de televisão |
| 10. Aspirador | 26. Sistema de aquecimento central | |
| 11. Máquina de lavar roupa | 27. Outro aparelho de aquecimento de ar | |
| 12. Máquina de secar roupa | | |
| 13. Máquina de lavar e secar roupa | | |
| 14. Máquina de lavar loiça | | |
| 15. Leitor de CD's | | |
| 16. Leitor de DVD's | | |

1.2 Metodologia estatística

1.2.1 Análise exploratória de dados

Com o objectivo de extrair indicadores elementares que caracterizassem os dados em estudo e que nos permitissem identificar, descrever e interpretar objectivamente algumas características relevantes das variáveis disponíveis, através de diferentes perspectivas, recorreremos à estatística descritiva.

Reconhecendo as capacidades de várias ferramentas, como Excel, SAS, SPSS, R, entre outros, para a análise estatística de dados, foram utilizadas nas diferentes fases da análise estatística e/ou modelação dos dados o SPSS e R. A escolha da utilização destas ferramentas teve origem na utilização das mesmas no decorrer das diferentes unidades curriculares do Mestrado.

Devido à dimensão das amostras, quer para Portugal, quer para Espanha, o SPSS demonstrou-se pouco robusto para tratamento e análise dos dados, apenas tendo sido utilizado para transformação e/ou codificação de algumas variáveis utilizadas na análise e obtenção de resultados referentes à análise preliminar dos dados.

1.2.2 Curva de Lorenz

Desenvolvida por Lorenz (1905) utilizámos a representação gráfica da distribuição das variáveis associadas ao rendimento e despesas dos agregados familiares, que ficou conhecida como curva de Lorenz.

Utilizando um acontecimento genérico, considere-se uma distribuição de frequências com classes (l_i, l_{i+1}) e frequências n_i , $i=1, 2, \dots, m$, e seja y_i o total da característica correspondente aos indivíduos da i -ésima classe. Estes valores y_i podem ser obtidos inicialmente através do cálculo do ponto médio $(n_i x_i)$ e de seguida, multiplicando o valor do ponto médio da classe pela frequência correspondente, $y_i \cong n_i x_i$. Quando os pontos médios não se distanciam muito dos valores médios de cada classe, os resultados são muito aproximados.

Calculamos de seguida as quantidades

$$p_i = \frac{\sum_{j=1}^i n_j}{\sum_{j=1}^m n_j} \quad \text{e} \quad q_i = \frac{\sum_{j=1}^i y_j}{\sum_{j=1}^m y_j},$$

Para $i=1, 2, \dots, m$ sabe-se que p_i representa a proporção de indivíduos que têm a característica com uma intensidade menor do que o limite superior da i -ésima classe, l_{i+1} ; q_i representa a proporção da totalidade da característica detida pelos mesmos.

Os valores p_i e q_i , $i=1, 2, \dots, m$, cumprem as relações.

$$p_i \geq q_i ; 0 \leq p_i \leq 1 ; 0 \leq q_i \leq 1$$

A recta de equação $y = x$ é designada por recta de igual distribuição e corresponde à situação que seria obtida quando $p_i = q_i$. Ao esboço da curva que se obtém unindo todos os pontos (p_i, q_i) chamamos curva de Lorenz. A concentração será maior quanto mais afastada estiver a curva da recta. Mais desenvolvimentos deste tema podem obter-se em Murteira *et al.* (1983).

1.2.3 Índice de GINI

De modo a medir a concentração do rendimento total pelos agregados da amostra, recorreu-se ao índice de Gini, que foi utilizado em 1912 pela primeira vez, por Corrado Gini e ficou conhecido como índice de concentração de Gini. Este é dado pela expressão:

$$IG = \frac{\sum_{i=1}^{n-1} (p_i - q_i)}{\sum_{i=1}^{n-1} p_i}$$

Repare-se que este índice é igual a zero quando $p_i = q_i$, ou seja, quando se verifica igual distribuição; igual a um quando $q_i = 0$, $i=1,2,\dots, n-1$, ou seja, quando a concentração é máxima; e aumenta com a diferença entre p_i e q_i , logo com a área entre a recta de igual concentração e a curva de Lorenz, ou seja, com o crescimento da concentração do atributo em estudo.

1.2.4 Modelos Lineares Generalizados (General Linear Models: GLM)

A construção de modelos para dados categóricos, surgiu em 1930. Nomes como Chester Bliss, R. A. Fisher, Frank Yates, J. Berkson, J. Cornfield ficarão para sempre associados a estes modelos. Mas é desde o início da década de '70 que se tem vindo a assistir a um acentuado impulso, na área de modelação estatística de regressão, em especial, nas áreas associadas às ciências biomédicas e sociais.

Em 1972, foram apresentados e propostos por Nelder e Wedderburn (1972), os modelos lineares generalizados, cuja ideia básica consiste em unificar o leque de opções para a distribuição da variável resposta, sendo que a mesma deve pertencer a uma distribuição dentro da família exponencial (Normal, Binomial, Poisson, Gama, Gaussiana inversa, ...). Foram os mesmos autores que propuseram um processo iterativo para a estimação dos parâmetros e incluíram o conceito de desvio, que tem vindo a ser utilizado na avaliação do ajustamento dos GLM, assim como no desenvolvimento de resíduos e medidas de diagnóstico.

A estrutura associada aos GLM tem três componentes:

1. Aleatória | identifica a variável resposta Y e selecciona uma distribuição de probabilidade.
2. Sistemática | que especifica as variáveis explicativas.
3. Função de ligação | especifica uma função $g(\cdot)$ que relaciona μ com o preditor linear

$$g(\mu) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

Em que x_1, \dots, x_n são as variáveis explicativas e β são os respectivos coeficientes.

De salientar que os GLM abrangem uma parte dos modelos mais comuns na análise estatística de dados, tal como ilustra a tabela 1.

Tabela 1: Alguns modelos lineares generalizados

Componente aleatória	Componente sistemática		Modelo
	f. ligação	covariáveis	
Normal	Identidade	Contínuas	Regressão linear
Normal	Identidade	Categorizadas	Análise de variância
Normal	Identidade	Mistas	Análise de covariância
Binomial	Logit	Mistas	Regressão logística
Poisson	Logarítmica	Mistas	Log-linear

Mas desde os anos '70 que têm sido publicados inúmeros trabalhos relacionados com o tema. Alguns dos resultados são apresentados em McCulloch e Searle (2001), McCulloch e Nelder (1989) e em muitos outros livros da literatura Estatística.

1.2.4.1 Regressão logística

1.2.4.1.1 Características do modelo

No modelo de regressão linear é assumido que a variável dependente Y é contínua e os resíduos aleatórios (e_i com $i = 1, \dots, n$) independentes e normalmente distribuídos com a mesma variância (σ^2). O modelo de regressão linear simples, com apenas uma variável independente, é dado por:

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

com $i = 1, \dots, n$ e e_i iid $N(0, \sigma^2)$

Os coeficientes do modelo (β_0 e β_1) são estimados pelo Método dos Mínimos Quadrados que consiste em encontrar a ordenada na origem e o declive da recta de forma a minimizar a soma dos quadrados dos resíduos aleatórios.

O modelo de regressão logística é um dos GLM mais utilizados para a análise de dados binários.

Quando se pretendem modelar dados e quando se utilizam muitas covariáveis é fundamental identificar o modelo mais parcimonioso e que proporcione uma boa interpretação do problema proposto e que o mesmo se ajuste bem aos dados que

possuímos. Para o cumprimento deste objectivo – selecção do melhor modelo – deve aliar-se o equilíbrio entre bom ajustamento, parcimónia e interpretação.

A aplicação dos GLM discretos para dados binários é efectuada usualmente com os modelos de regressão logística, probit e complementar log-log, dependendo se a função de distribuição for a logística, normal reduzida ou de Gumbel, respectivamente. A utilização de uma ou outra função de ligação e conseqüentemente a opção pelo modelo a utilizar depende da situação em causa. De um modo geral, os modelos probit e logístico possuem capacidade de adaptação muito semelhante, contudo o modelo probit é menos flexível, pois não pode ser aplicado directamente com mais de uma variável (<http://www.ats.ucla.edu>). O modelo log-log, devido à função complementar log-log, pode dar respostas diferentes, sendo mais utilizado para análise de dados relativos a incidências de doenças. Sobre os dois últimos modelos podem ser consultados mais desenvolvimentos em Turkman *et al.* (2000) ou (<http://www.ats.ucla.edu>).

Tal como referido anteriormente, os dados binários são a forma mais comum de dados categorizados, sendo a regressão logística o método de modelação estatística de dados binários mais popular. No entanto, mesmo quando a variável de interesse não é originalmente do tipo binário, alguns investigadores optam por dicotomizar a variável, de modo a que a probabilidade de sucesso possa ser modelada através da regressão logística (Paula, 2004). As variáveis de que dispomos para proceder selecção do modelo mais adequado, resultam igualmente de um processo de categorização, tal como apresentado na Tabela 5. Por esse motivo, será abordada com maior detalhe a metodologia associada e este tipo de dados.

Uma das características que distingue um modelo de regressão logística de um modelo de regressão linear é que a variável resposta no primeiro modelo é binária e assumindo os valores 1 e 0, conforme ocorre ou não o evento em estudo, de acordo com uma sequência de eventos de Bernoulli. Considerando Y_i a variável resposta, X_i a variável explicativa e β_0 e β_1 os coeficientes do modelo linear tem-se que o valor médio é dado por

$$\begin{cases} Y_i = 1 \rightarrow P(Y_i = 1) = \pi_i \\ Y_i = 0 \rightarrow P(Y_i = 0) = 1 - \pi_i \end{cases}$$

$$e \ E(Y_i) = \beta_0 + \beta_1 X_i = \pi_i$$

As variâncias, para o modelo apresentado são dadas por

$$\sigma^2(Y_i) = E\left[(Y_i - E(Y_i))^2\right] = (1 - \pi_i)^2 \pi_i + (0 - \pi_i)^2 (1 - \pi_i)$$

$$E(Y_i)(1 - E(Y_i)) = \sigma^2(Y_i) = \pi_i(1 - \pi_i).$$

Ou seja, estamos perante variâncias heterogêneas. Ora, como os resíduos são dados por

$$\varepsilon_i = Y_i - \pi_i, \text{ (com } Y_i \text{ constante),}$$

as variâncias dos resíduos são dadas por

$$\sigma^2(\varepsilon_i) = \pi_i(1 - \pi_i) = (\beta_0 + \beta_1 X_i)(1 - \beta_0 - \beta_1 X_i) \quad \text{depende de } X_i$$

Deste modo verificamos, que não é válido nenhum dos pressupostos inerentes ao modelo de regressão linear.

Por outro lado, observe-se que o valor médio da variável resposta está limitado entre 0 e 1

$$0 \leq E(Y \mid x) = \pi \leq 1,$$

e, conseqüentemente, esta restrição é inapropriada para uma função de resposta linear.

Neste tipo de resposta a forma da função tem tendência a ser curvilínea. A função logística tem uma curva da resposta em forma de S quando a ocorrência de eventos aumenta com o valor da variável explicativa, fazendo lembrar uma representação gráfica de uma distribuição cumulativa de uma variável aleatória. Por outro lado, a curva da resposta revela uma forma de S invertido quando a ocorrência de eventos diminui com o valor da variável explicativa.

A resposta pode ser linearizada através da função logit, pois

$$\text{logit } \pi(x) = \log\left[\frac{\pi(x)}{1 - \pi(x)}\right] = \beta_0 + \beta_1 x \text{ e assim}$$

$$E(Y) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \text{ ou } E(Y) = [1 + \exp(-\beta_0 - \beta_1 X)]^{-1}$$

Esta transformação assume uma grande importância, pois é linear nos seus parâmetros, pode ser contínuo e, dependendo da amplitude de x pode variar entre $-\infty$ e $+\infty$. (Hosmer e Lemeshow, 2000).

A grande popularidade da regressão logística assenta fundamentalmente na facilidade de interpretação dos parâmetros do modelo que a função logit permite, uma vez que o quociente entre a probabilidade de ocorrência do evento e a probabilidade de não ocorrência traduzem o odds (chance) de ocorrer o sucesso relativamente ao insucesso.

1.2.4.1.2. Interpretação dos coeficientes

Sendo

$$\text{logit}\pi(X_j) = \hat{\beta}_0 + \hat{\beta}_j X_j$$

tem-se

$$\text{logit}\pi(X_j + 1) - \text{logit}\pi(X_j) = \hat{\beta}_j$$

logo

$$\log_e(\text{odds}_2) - \log_e(\text{odds}_1) = \log_e\left(\frac{\text{odds}_2}{\text{odds}_1}\right) = \hat{\beta}_j$$

Consequentemente

$$\text{OR} = \frac{\text{odds}_2}{\text{odds}_1} = \exp(\hat{\beta}_j)$$

A exponencial do coeficiente representa a razão de chances ou possibilidades (*Odds Ratio*) da ocorrência do evento relativamente à não ocorrência quando a variável independente X_j aumenta uma unidade (ou está num dada categoria, no caso de ser categórica) relativamente às chances ou possibilidades da ocorrência do evento quando X_j permanece constante (está na categoria de referência). Ou seja, quando X_j varia uma unidade as chances de obter o sucesso aumentam (caso $\text{OR} > 1$) ou diminuem (Caso $\text{OR} < 1$) $\hat{\beta}_j$ unidades.

Para valores de n grandes, os intervalos de confiança, mais usuais para os parâmetros de regressão são os de Wald e são definidos pela expressão (Agresti, 2007):

$$\hat{\beta}_i \pm Z_{1-\alpha/2} \hat{\sigma}_{\hat{\beta}_i}$$

Com a estrutura de dados de que dispomos nenhum dos softwares utilizados calcula o intervalo de confiança, por isso é necessário recorrer à matriz de covariância dos parâmetros do modelo quando temos interações significativas e a partir desta calcular a covariância associada aos coeficientes intervenientes como se pode ver em Hosmer e Lemeshow (2000).

1.2.4.1.3 Avaliação da bondade do ajustamento do modelo

À medida que são incluídas as variáveis independentes no modelo, interessa saber se conseguimos obter informações acerca da variável explicativa com ou sem a variável em causa. Essa questão pode ser respondida através da comparação dos valores observados da variável resposta com os valores esperados de cada um dos modelos, o primeiro com a variável e o segundo sem a mesma. É, neste caso, conveniente considerar que o valor observado da variável resposta é também um valor esperado resultante de um modelo saturado (e.g. que contem todas as variáveis independentes), a comparação de valores observados com valores esperados é baseada na expressão:

$$D = -2 \ln \left(\frac{\text{verosimilhança do modelo ajustado}}{\text{verosimilhança do modelo saturado}} \right) \quad (1)$$

A quantidade dentro do parêntesis da expressão (1) é a razão de verosimilhança.

Como no modelo saturado $\hat{\pi}_i = y_i$

$$D = -2 \ln(\text{verosimilhança do modelo ajustado}) \quad \text{ou} \quad D = -2LL$$

O $-2LL$ é um indicador da mediocridade do ajustamento do modelo aos dados, ou seja, quanto maior for este resultado, pior é o ajustamento.

Com o objectivo de verificar a significância da variável independente deve ser comparado o modelo com e sem a mesma, sendo o cálculo da estatística de teste para avaliar a significância do modelo G a diferença entre o $-2LL$ dos dois modelos (diferença entre o valor da verosimilhança do modelo final e do modelo reduzido),

$$\begin{aligned} G &= D(\text{modelo sem variavel}) - D(\text{modelo com variavel}) = \\ &= -2 \ln \left(\frac{\text{verosimilhança sem variavel}}{\text{verosimilhança com variavel}} \right) \end{aligned}$$

Permitindo não rejeitar ou rejeitar a hipótese nula de que o modelo não é significativo.

A forma mais usual de testar se uma dada variável independente é significativa para o modelo consiste em recorrer ao Teste de Wald, o qual está implementado em todos os softwares. A estatística de teste é dada pela expressão

$$T_{Wald_i} = \frac{\hat{\beta}_i}{\hat{\sigma}(\hat{\beta}_i)} \dot{\sim} N(0,1)$$

onde $\hat{\beta}_i$ é o estimador do coeficiente β_i e $\hat{\sigma}(\hat{\beta}_i)$ é o estimador do erro padrão de $\hat{\beta}_i$. Esta estatística tem distribuição t -student, que se aproxima assintoticamente da

distribuição $N(0,1)$, para amostras grandes. Hauck *et al.* (1977), ao estudarem o desempenho do teste de Wald, concluíram que o mesmo era conservativo, pelo que as conclusões retiradas com este teste devem ser efectuadas com prudência.

Para testar a significância do ajustamento do modelo completo, utiliza-se a estatística de teste do Qui-quadrado de Pearson (Hosmer & Lemeshow, 2000):

$$\chi_p^2 = \sum_{j=1}^J \frac{(O_j - E_j)^2}{E_j},$$

O_j e E_j são, respectivamente, o número de sucessos observados e de sucessos esperados para a célula j . Outros autores (McCullag & Nelder, 1989) utilizam a estatística *Deviance*:

$$D = -2Ln \left[\frac{L_c}{L_s} \right],$$

onde L_c é a verosimilhança do modelo ajustado e L_s é a verosimilhança do modelo com todas as variáveis independentes. A significância dos testes do Qui-quadrado e da Deviance permitem rejeitar ou não a hipótese “o modelo ajusta-se aos dados”.

Estes testes de bondade de ajustamento só fazem sentido para os dados agrupados. Quando calculadas para os modelos de regressão logística com preditores contínuos estas estatísticas não têm distribuição aproximadamente qui-quadrado. Esta questão é ultrapassada, criando categorias para cada variável explicativa e aplicar essas estatísticas às contagens observadas e ajustadas em cada categoria.

O melhor procedimento, segundo mostraram Hosmer e Lemeshow (2000) é obter os intervalos de categorização de acordo com as probabilidades estimadas, sendo formados 10 grupos de igual tamanho, em que o primeiro se refere ao decil das observações com maiores probabilidades estimadas e assim sucessivamente. O teste de bondade de ajustamento de Hosmer-Lemeshow usa uma estatística de teste de Pearson para comparar as contagens observadas com as estimadas pelo modelo de regressão logística.

Os modelos *logit* não têm uma boa medida da qualidade do ajustamento intuitiva como o coeficiente de determinação para os modelos lineares.

A estimativa dos pseudo R^2 de Cox & Snell e Nagelkerke foram propostas para colmatar esta falta, mas Hosmer e Lemeshow (2000) referem que os pseudo R^2 não são medidas da variabilidade explicada pelo modelo, pois são baseados na comparação do modelo ajustado com o modelo nulo.

O critério Akaike information criterion (AIC) permite avaliar a bondade de ajustamento do modelo, através dos valores ajustados. Ou seja, verifica a distância que separa os valores ajustados dos valores esperados. O melhor modelo é aquele que apresenta os valores ajustados mais próximos dos valores reais, sendo o que apresenta um menor valor de AIC.

Este critério é baseado na função log-verosimilhança, com introdução de um factor de correcção como modo de penalização da complexidade do modelo e é definido como

$$AIC = -2(\log \text{verosimilhança} - n^\circ \text{ de parâmetros no modelo})$$

1.2.4.1.4 Selecção de variáveis

Até agora foram focados procedimentos associados à inclusão de variáveis, significância e interpretação dos coeficientes dos modelos de regressão logístico. Mas não deve ser indiferente o facto de que a inclusão de uma determinada variável pode depender do objectivo do estudo em causa, sem que o modelo final possa vir a ser o mais parcimonioso. Em estudos epidemiológicos é sugerida a inclusão de variáveis que sejam consideradas relevantes do ponto de vista clínico e/ou intuitivo, independentemente da sua significância estatística, contudo é necessário assegurar que o modelo obtido não seja demasiado denso o que certamente poderá provocar um grande número de coeficientes estimados, resultando em modelos pouco realistas. Mais desenvolvimentos sobre esta abordagem podem ser consultados em Harrell *et al* (1996).

Apesar de nos softwares estatísticos estarem implementados vários métodos automáticos de selecção de variáveis (*backward*, *forward* e *stepwise*, entre outros), estes devem servir como ajuda e não como a forma de definir sem reservas o modelo final. De modo a assegurar a selecção das variáveis de forma criteriosa podem ser seguidos alguns passos, sendo este processo, de uma forma geral, parecido com os critérios utilizados na regressão linear:

- O processo de selecção deve iniciar-se com a análise da influência de cada variável relativamente à variável resposta.
- Qualquer variável que, no ponto anterior, obtenha um *p-value* < 0.25 é candidata ao modelo multivariado, juntamente com as variáveis que se considerem importantes ainda que tenham obtido significância. De seguida, a análise deve prosseguir com todas as variáveis seleccionadas.

- Seguidas as etapas descritas, deve ser avaliada a importância de cada variável incluída no modelo obtido.
- Procedemos, então, à eliminação de cada variável não significativa no modelo por ordem decrescente dos *p-values* associados e até que no modelo fiquem apenas variáveis significativas a um nível de 5% (ou mesmo 10% como é aconselhado por muitos autores).
- Deve, contudo, ser salientado que o processo de incluir, retirar, reajustar e verificar deve ser repetido até que se considere que todas as variáveis importantes estão incluídas no modelo. Este procedimento está associado à análise preliminar de efeitos principais do modelo.
- Após a obtenção do modelo no qual é assumido que estão incluídas as variáveis essenciais, deve ser dado início a uma análise mais atenta às mesmas. Neste patamar está incluída a verificação de linearidade do logit com as variáveis contínuas, no caso de existirem. Assumir a linearidade no logit na fase de selecção das variáveis é comum sendo consistente com o objectivo de determinar se essa variável deve ou não ser incluída no modelo. Usualmente faz-se uma representação gráfica dos pontos médios dos intervalos obtidos pela categorização da variável contínua usando os quartis como ponto de corte contra os coeficientes de cada categoria (a de referência naturalmente com o valor nulo) do modelo ajustado substituindo a variável contínua pela variável contínua categorizada. Naturalmente, espera-se que desta representação gráfica resulte aproximadamente uma recta.
- Depois de verificados os efeitos principais do modelo, devem ser testadas as interacções entre as variáveis do modelo. Em qualquer modelo uma interacção entre duas variáveis implica que o efeito de uma das variáveis não é constante sobre os níveis da outra variável, começando por verificar as interacções significativas uma a uma e posteriormente quais destas são significativas no modelo.

Porém, antes que qualquer modelo seja utilizado para inferência dos resultados obtidos, devem ser verificadas a sua adequação e ajustamento. Uma avaliação completa do modelo ajustado envolve o cálculo das medidas sumárias da distância entre valores observados e estimados (bondade do ajustamento do modelo) e uma análise gráfica de resíduos com o objectivo de procurar observações individuais influentes sobre a estimação dos parâmetros do modelo ou *outliers*.

1.2.4.1.5 Diagnóstico do Modelo

A análise de resíduos é importante não só para a avaliação local da qualidade de ajustamento de um modelo como para a identificação de observações mal ajustadas, ou seja, para as quais não se conseguem obter boas explicações através do modelo.

Informalmente pode proceder-se a uma análise aos resíduos através de representações gráficas adequadas.

Os resíduos de Pearson standardizados são utilizados para medir a diferença entre os valores observados e os valores ajustados.

$$R_i^P = \frac{y_i - n_i \hat{\pi}_i}{\sqrt{n_i \hat{\pi}_i (1 - \hat{\pi}_i) (1 - h_i)}}$$

Contudo apresentam uma desvantagem relacionada com a distribuição que lhe está associada, pois é, de um modo geral, bastante assimétrica para modelos não normais.

Em vez destes resíduos costumam utilizar-se os resíduos da função *Deviance*, onde o desvio residual corresponde à i-ésima observação para a função desvio é dado por

$$R_i^D = \delta_i [y_i \log_e(\hat{\pi}_i) + (1 - y_i) \log_e(1 - \hat{\pi}_i)], \quad \delta_i = \text{ sinal}(y_i - \hat{\pi}_i)$$

onde h_{ii} são os elementos diagonais da matriz Hessiana (Hosmer e Lemeshow (2000)). Estes resíduos são usualmente representados contra os indivíduos ou contra o preditor linear, sendo espectável que se distribuam aleatoriamente em torno da origem.

Os resíduos parciais permitem apurar a adequabilidade da escala em que as covariáveis contínuas estão representadas, sendo usualmente representados contra os valores observados das covariáveis. Estes podem ser definidos como

$$R_{parciais} = u - \hat{\eta} + \hat{\gamma}x$$

onde $u - \hat{\eta}$ é o vector dos resíduos medidos na escala linear, u é a variável dependente ajustada já definida e $\hat{\gamma}$ é a estimativa do parâmetro para a variável explicativa em consideração. Se a escala de x é satisfatória, então os valores do gráfico devem ser aproximadamente lineares.

Quando se tenta detectar a existência de uma ou mais observações mal ajustadas pelo modelo, o objectivo passa não só por identificar as mesmas como por medir o impacto da sua existência no modelo. Uma observação é influente se a sua exclusão do modelo produzir alterações significativas nas estimativas dos parâmetros do modelo. Por este motivo, a sua presença pode dar origem a um impacto indevido nas

conclusões a retirar do modelo. Este tipo de observações não tem, necessariamente, resíduos elevados. Após a identificação dos casos mais influentes pode obter-se exactamente a influência da sua omissão. A medida utilizada para identificar as observações influentes é a distância de Cook,

$$D_i = \frac{(\hat{\beta} - \hat{\beta}_{(i)})^T (Z^T W Z) (\hat{\beta} - \hat{\beta}_{(i)})}{p \hat{\phi}}$$

em que $\hat{\beta}_{(i)}$ e $\hat{\beta}$ representam respectivamente as estimativas de máxima verosimilhança do vector parâmetro β obtidas da amostra sem a observação (y_i, z_i) e da amostra com todas as observações.

A influência de cada observação na estimação de cada um dos coeficientes de regressão pode ser estimada pelos *DfBet*s. Os *DfBet*s podem ser calculados pela expressão:

$$DfBeta_{ij} = \hat{\beta}_i - \hat{\beta}_{i(-j)}$$

onde $\hat{\beta}_i$ representa a estimativa do coeficiente de regressão ajustado para todas as observações e $\hat{\beta}_{i(-j)}$ a estimativa do coeficiente de regressão ajustado sem a observação j .

Os resultados da regressão logística podem ser usados para efeitos de classificação da resposta: caso a probabilidade do acontecimento seja maior ou igual a um determinado valor (ponto de corte), o modelo prediz que o evento vai ocorrer.

Tabelas de classificação são formas intuitivas de sistematizar resultados de ajustamento de um modelo de regressão logístico. Mas a precisão ou não precisão das mesmas não representam um critério de ajustamento, embora possam ser úteis quando utilizadas com outras medidas, baseadas nos resíduos. Medidas como sensibilidade e especificidade derivam de tabelas de classificação.

Os valores afectos à sensibilidade correspondem à percentagem de ocorrências correctamente preditas. Sendo a percentagem de não ocorrências correctamente preditas o valor dado pela especificidade.

Inicialmente as expressões sensibilidade e especificidade foram utilizadas para descrever procedimentos associados a testes para doenças. Quando é efectuado apenas um teste, a pessoa pode ter, de facto, a doença ou não ter a doença. O resultado do teste efectuado ao paciente pode ser positivo, indicando a presença da doença ou o teste pode dar resultado negativo, indicando a ausência da mesma. Podemos verificar o resultado do teste em coluna e o estado real da pessoa que efectuou o teste, em linha.

		Resultado do teste (T)	
		Positivo (+)	Negativo (-)
Estado real (S)	Doença (+)	a	b
	Não doença (-)	c	d

Sensibilidade, pode ser definida como a probabilidade de um teste diagnosticar que uma pessoa tem uma doença quando de facto tem. É dada por $P(T^+|S^+) = \frac{a}{a+b}$.

A probabilidade de, por exemplo, um teste não diagnosticar a doença quando uma pessoa de facto não a tem é atribuída à especificidade e é dada por $P(T^-|S^-) = \frac{d}{c+d}$.

Idealmente, os testes aplicados aos pacientes deveriam ter associado um elevado valor destas duas medidas. Mas é bom não subavaliar falsos positivos e falso negativos.

O falso positivo ocorre quando o teste indica resultado positivo para uma pessoa que não tem essa doença. A taxa de falsos positivos é dada por $P(S^-|T^+) = \frac{c}{a+c}$.

O falso negativo ocorre quando o teste indica resultado negativo para uma pessoa que realmente tem a doença. A taxa de falsos negativos é dada por $P(S^+|T^-) = \frac{b}{b+d}$.

A Curva ROC (*receiver operating characteristic*) é um gráfico que representa a sensibilidade como uma função de 1-especificidade, para vários pontos de corte, que são obtidos através das probabilidades estimadas pelo modelo de regressão logística. O indicador resultante da curva é mais informativo que a tabela de classificação, pois resume o poder preditivo para um possível valor π_0 . Quando π_0 chega próximo de 0, praticamente todas as predições são $\hat{y}=1$, então a sensibilidade está perto de 1 e especificidade está perto de 0 e o ponto para (1-especificidade, sensibilidade) tem coordenadas perto de (1,1).

Para uma dada especificidade, o melhor poder preditivo corresponde a uma sensibilidade alta.

Embora o melhor ponto de corte dependa de um contexto específico é muitas vezes usual tomar como ponto de corte o valor que torna máxima a sensibilidade e a especificidade.

A área abaixo da curva é idêntica ao valor da medida do poder preditivo, chamado índice de concordância, que tem por objectivo estimar a probabilidade das previsões e dos resultados serem concordantes.

A área abaixo da curva fornece uma medida da habilidade do modelo discriminar entre indivíduos que têm atributo e os que não o têm, varia entre 0 e 1.

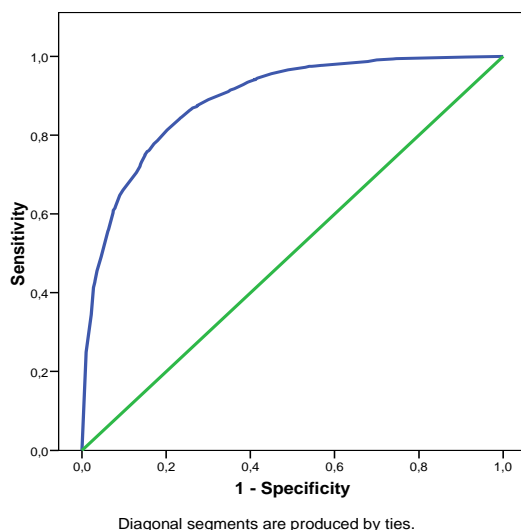


Figura 1: Exemplo de curva ROC

Os valores na Tabela 2 permitem avaliar o poder discriminativo da área sob a curva de ROC, (Hosmer e Lemeshow, 2000).

Tabela 2: Poder discriminativo da área sob a curva de ROC

Área sob a curva (A)	Diagnóstico (poder discriminativo)
0.5	Modelo sem poder discriminativo
$0.6 \leq A < 0.7$	Pobre
$0.7 \leq A < 0.8$	Aceitável
$0.8 \leq A < 0.9$	Excelente
$0.9 \leq A < 1$	Excepcional

Capítulo II

2.1 Análise dos dados

2.1.1 Estatística Descritiva

Após a descrição gráfica importa proceder à leitura das características numéricas obtidas através das medidas de tendência central, localização, dispersão, assimetria e concentração.

O facto de a média ser calculada a partir de todas as observações torna-a sensível à existência de valores extremos, podendo fazê-la deslocar-se para o lado mais afastado, não permitindo que o valor apresentado caracterize correctamente o conjunto de dados. Quando estes casos se verificam a melhor opção é utilizar outra medida de localização que não seja afectada pelos valores extremos.

A mediana é a medida mais adequada para descrever conjuntos de dados que incluam os valores extremos, pois é calculada com base na posição central na série ordenada das observações, ou seja, dispondo as observações por ordem crescente, a mediana é o valor central que divide a distribuição para que, pelo menos 50% das observações sejam menores ou iguais a esse valor.

Na Tabela 3 podemos verificar os resultados obtidos para as variáveis associadas às despesas em ambos os países, relativamente às medidas amostrais.

Através dos resultados obtidos, podemos verificar que a médias das despesas total e monetária são mais elevadas em Espanha e que, apesar das variáveis apresentarem amplitudes de valores muito elevadas, o que deverá, influenciar a leitura através da média, podemos evidenciar a mesma conclusão, através dos valores da mediana. É em Espanha que se verificam os maiores desvios relativamente à média e que através do coeficiente de variação podemos verificar que se situam nos 96%, no caso da despesa total e não monetária e nos 104%, na despesa monetária. Portugal apresenta a variação mais elevada na despesa monetária, com 96%.

Tabela 3: Medidas amostrais relativas às variáveis sobre despesa, em Portugal e Espanha

Medidas amostrais	Despesa total		Despesa monetária		Despesa não monetária	
	PT	ES	PT	ES	PT	ES
Média	16 184	19 891	12 027	16 249	4 157	3 646
IC a 95% p/ a Média – Limite Inferior	15 949	19 624	11 826	16 011	4 086	3 597
IC a 95% p/ a Média – Limite Superior	16 419	20 159	12 229	16 488	4 227	3 696
Mediana	13 074	13 743	9 221	10 082	3 600	3 148
Variância	149 988 920	361 791 256	109 773 519	288 293 128	13 391 872	12 231 716
Desvio-Padrão	12 247	19 021	10 477	16 979	3 659	3 497
Coefficiente de Variação	76%	96%	96%	104%	88%	96%
Mínimo	108	3	108	3	0	0
Máximo	196 509	209 020	150 917	209 020	53 366	50 030
Amplitude	196 401	209 017	150 809	209 017	53 366	50 030
Amplitude Interquartil	12 130	22 973	10 163	19 113	3 832	4 573
Assimetria	2,52	1,91	2,63	2,15	2,65	2,05
Erro-Padrão (Assimetria)	0,02	0,02	0,02	0,02	0,02	0,02
Curtose	12,69	6,11	12,59	7,74	15,85	10,08
Erro-Padrão (Curtose)	0,05	0,04	0,05	0,04	0,05	0,04

Quanto à assimetria, os histogramas obtidos e representados na Figura 2 revelam estruturas assimétricas, representando a forma da distribuição das despesas, denotando maior concentração das despesas na zona de valores iniciais e reflectem os resultados obtidos na Tabela 3, em que se verificam valores positivos elevados para o quociente entre o coeficiente de assimetria e o seu desvio padrão, pelo que estamos perante um enviesamento à direita ou assimetria positiva.

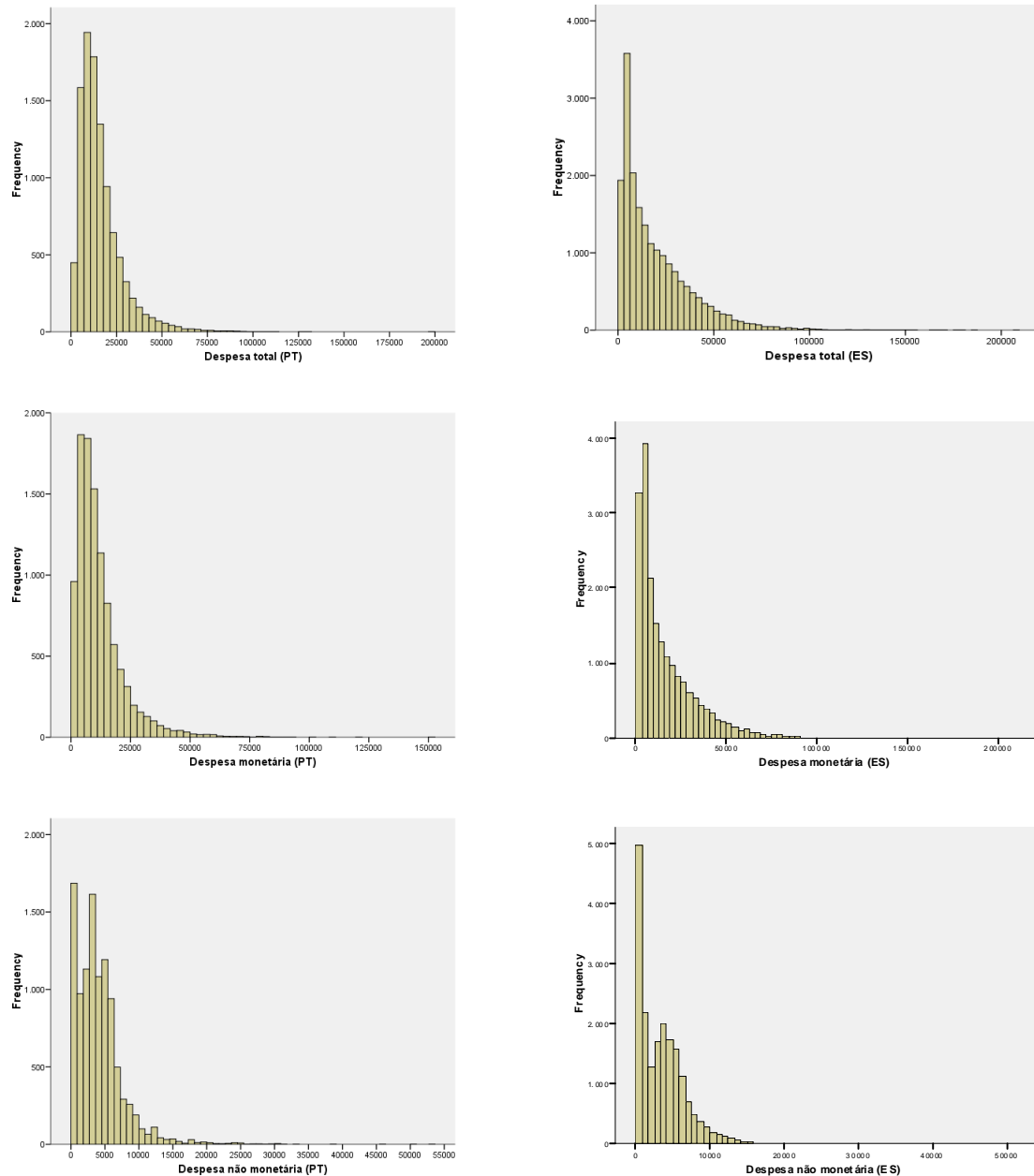


Figura 2: Histogramas das variáveis despesa total, despesa monetária e despesa não monetária, em Portugal e Espanha

Em relação ao grau de achatamento, podemos verificar através dos resultados obtidos na Tabela 3 (valores muito elevados para os quocientes entre o coeficiente de curtose e o seu desvio padrão) e observando a Figura 2, que todas as distribuições são do tipo leptocúrtico.

Através dos valores standardizados, representados nas caixas de bigodes (Figura 3), pode verificar-se graficamente a dispersão dos mesmos. Para todas as variáveis observadas podem ser facilmente identificadas muitas observações extremas, outliers moderados e severos. São as observações 3529 e 6960, em Portugal e Espanha,

respectivamente que mais se destacam nas variáveis despesa total e despesa monetária. No entanto, dada a elevada amplitude dos valores observados, mesmo com as variáveis tipificadas se torna muito difícil comparar as caixas de bigodes para os dois países.

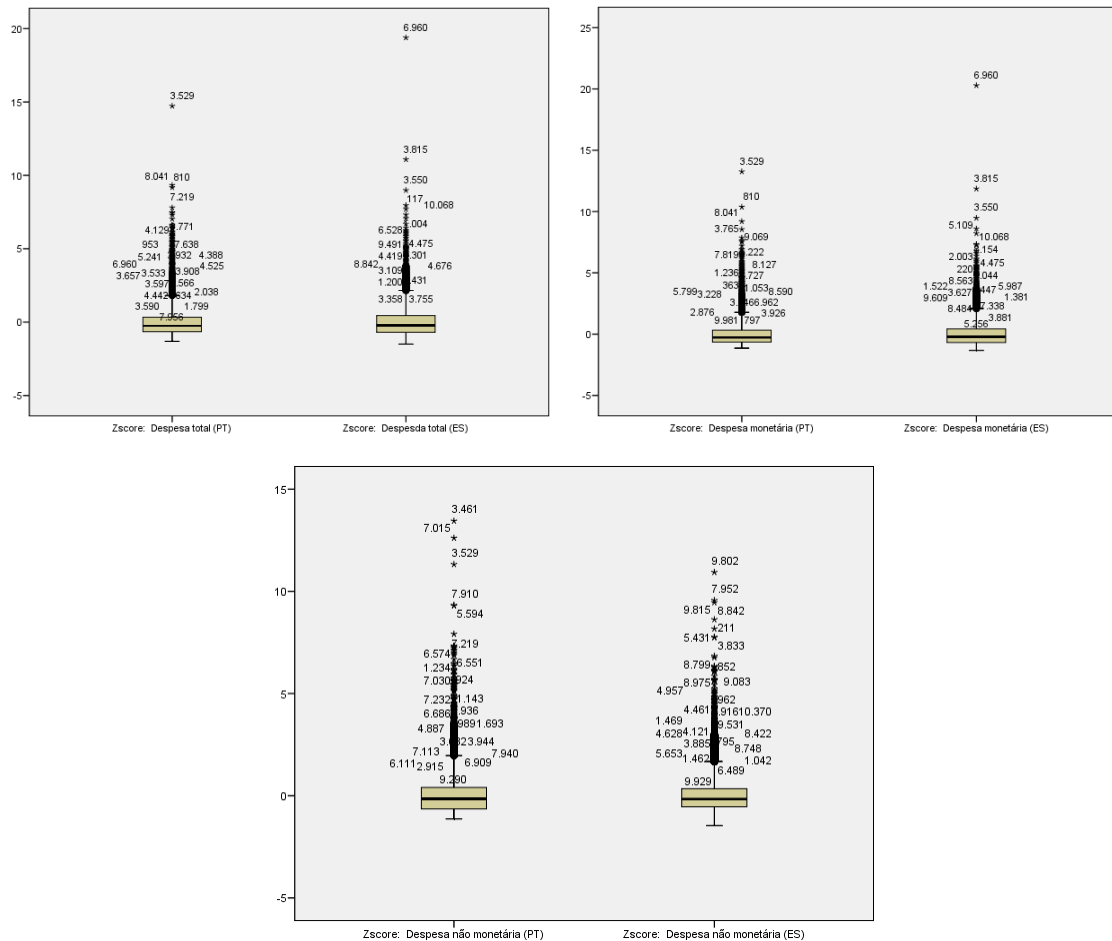


Figura 3: Caixas de bigodes das variáveis despesa total, despesa monetária e despesa não monetária, em Portugal e Espanha

A Tabela 4 inclui os resultados obtidos para as variáveis associadas ao rendimento em ambos os países, relativamente às medidas amostrais.

Tabela 4: Medidas amostrais relativas às variáveis sobre rendimentos, em Portugal e Espanha

Medidas amostrais	Rendimento total		Rendimento monetário	Rendimento não monetário
	PT	ES	PT	PT
Média	20 615	25 338	16 459	4 157
IC a 95% p/ a Média – Limite Inferior	20 278	25 101	16 152	4 086
IC a 95% p/ a Média – Limite Superior	20 952	25 574	16 766	4 227
Mediana	16 511	22 134	12 600	3 600
Variância	307 512 663	283 123 452	255 256 494	13 391 872
Desvio-Padrão	17 536	16 826	15 977	3 659
Coefficiente de Variação	85%	66%	97%	88%
Mínimo	600	0	500	0
Máximo	460 690	280 000	460 690	53 366
Amplitude	460 090	280 000	460 190	53 366
Amplitude Interquartil	15 331	18 200	13 248	3 832
Assimetria	5,32	2,53	6,27	2,65
Erro-Padrão (Assimetria)	0,02	0,02	0,02	0,02
Curtose	72,87	16,56	100,20	15,85
Erro-Padrão (Curtose)	0,05	0,04	0,05	0,05

A média do rendimento total é mais elevada em Espanha, sendo Portugal que apresenta a maior variabilidade (Coeficiente de variação de 85% contra o valor de 66% obtido para a Espanha). Contudo, ao contrário dos resultados obtidos para as variáveis relacionadas com as despesas, é Portugal que apresenta uma maior amplitude de valores relativamente ao rendimento total.

Para as variáveis cujos dados apenas estão disponíveis em Portugal, podemos verificar que o rendimento não monetário tem um peso muito inferior no rendimento total, pois o valor máximo nesta variável é de apenas 53366 €, face aos 460090 € relativos ao rendimento total.

Os histogramas das variáveis em estudo (Figura 4), representam uma assimetria positiva pela forma como os rendimentos estão distribuídos. Neste caso os valores da curtose são muito mais elevados, indicando distribuições com os valores menos concentrados em torno da média, mais ou menos um desvio padrão, caudas longas e pesadas, maior probabilidade de ter valores próximos à média e de ter valores extremos.

Os valores standardizados referentes ao rendimento estão representados nas caixas de bigodes da Figura 5. Estes demonstram não só a dispersão dos valores, como também as observações mais extremas que, no caso de Portugal, salientamos as observações 3765, 8103 e 2329, por mais se afastarem dos restantes e serem comuns ao rendimento total e rendimento monetário. A pequena dimensão das caixas e dos respectivos bigodes é influenciada pela reduzida amplitude interquartil e pela elevada amplitude dos rendimentos dos agregados.

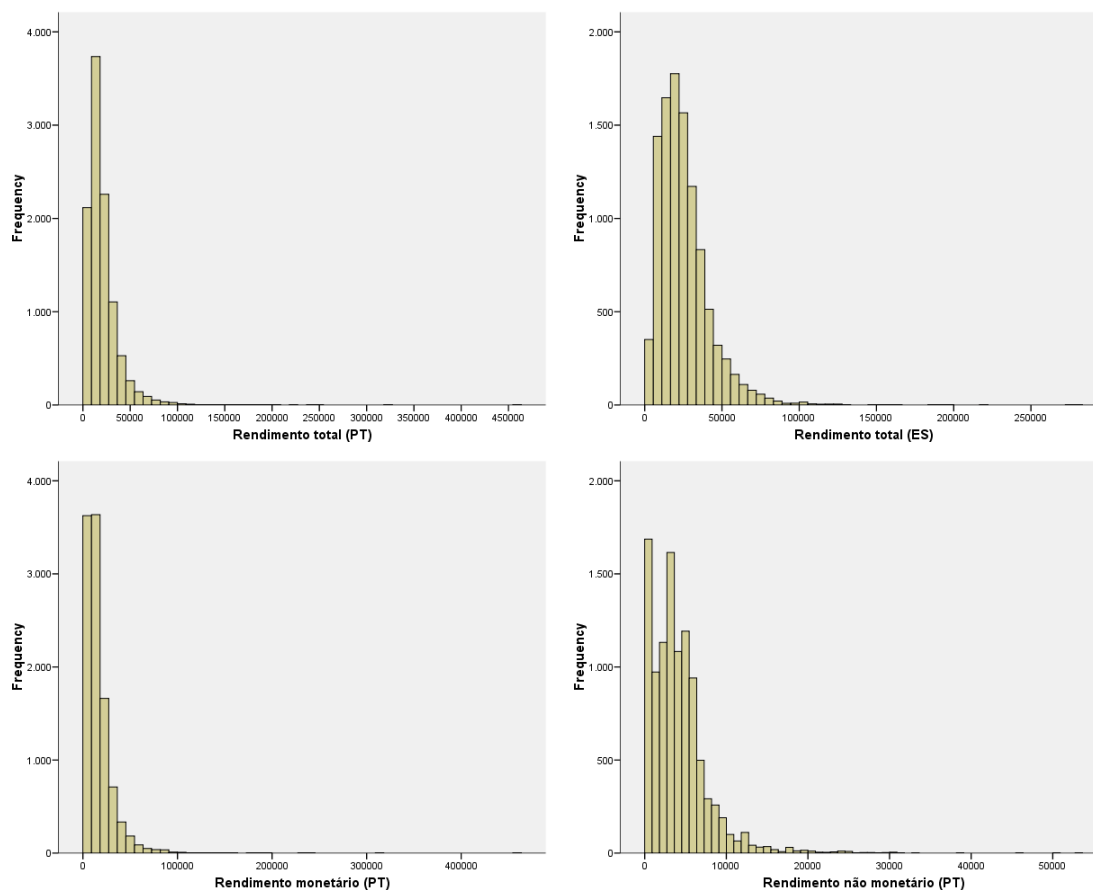


Figura 4: Histogramas das variáveis rendimento total, em Portugal e Espanha e despesa monetária e despesa não monetária, em Portugal

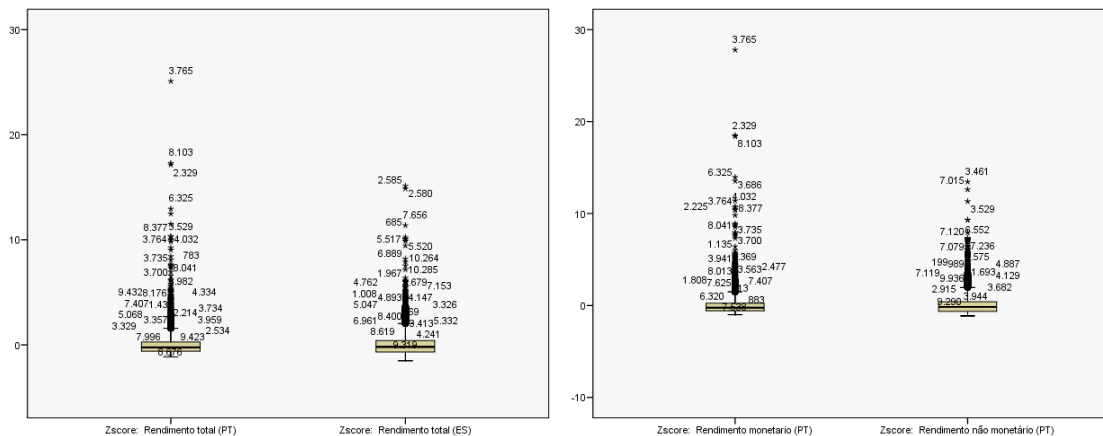


Figura 5: Caixas de bigodes das variáveis rendimento total, em Portugal e Espanha e despesa monetária e despesa não monetária, em Portugal

2.1.2 Estudo Gráfico da Distribuição das Variáveis

Ainda durante a análise exploratória dos dados, foram utilizados QQ-plots (gráficos Quartil) para verificar se a variável resposta poderia ser modelada por alguma distribuição conhecida e, em particular, por alguma distribuição pertencendo à família da exponencial, com o objectivo de mais tarde a podermos modelar usando a teoria dos modelos lineares generalizados.

Foram utilizadas várias distribuições pertencentes à família exponencial mas, através dos gráficos, verificámos que os dados não apresentavam o comportamento linear que se deve verificar quando se encontra a distribuição correcta ou pelo menos aproximada. Na Figura 6 ilustramos os resultados obtidos para 2 distribuições.

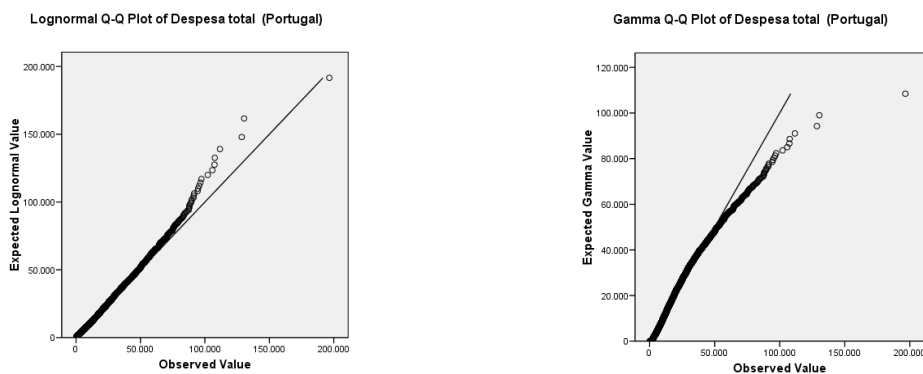


Figura 6: Q-Q plot das distribuições Lognormal e Gama, variável despesa total (Portugal)

Várias foram as transformações que também se tentaram, mas sem sucesso. Depois de verificada a impossibilidade de modelar a distribuição da variável resposta (despesa total) e de outras através de uma distribuição conhecida, optou-se pela categorização das mesmas, de forma a definir um percurso na análise e modelação dos dados.

As vantagens associadas à categorização estão relacionadas com o objectivo de cada estudo. Neste caso, optou-se pela categorização das variáveis identificando o nº de casos acima e abaixo da média ou acima e abaixo da mediana. Através da categorização podemos definir como objectivo encontrar um modelo que explique as variações da variável resposta (despesa total dos agregados familiares), de acordo com diferentes níveis de variáveis explicativas. As variáveis apresentadas na Tabela 5 representam as codificações utilizadas em ambos os países.

2.1.3 Classes de variáveis

Para além das codificações apresentadas anteriormente prosseguimos a análise das variáveis como despesas e rendimentos agrupando os dados em classes.

Este método de representação permitiu, entre outros, verificar e quantificar o nível de concentração destas variáveis, tornando possível verificar se, por exemplo, o rendimento total está igualmente repartido pelos agregados ou se existe uma maior concentração sobre uma ou mais classes e, se os níveis de concentração apresentam comportamentos semelhantes em Portugal e em Espanha.

Tabela 5: Tabela de codificação das variáveis referentes aos dados de Portugal e Espanha

Variável	Descrição	Codificação
DespTot_MEDIA	Despesa Total	0=abaixo da média 1=acima da média
DespTot_MEDIANA		0=abaixo da mediana 1=acima a mediana
DespTot_Q		0=Q1/4 1=Q2/4 2=Q3/4 3=Q4/4
RendTot_MEDIA	Rendimento Total	0=abaixo da média 1=acima da média
RendTot_MEDIANA		0=abaixo da mediana 1=acima a mediana
RendTot_Q		0=Q1/4 1=Q2/4 2=Q3/4 3=Q4/4
Gurb	Grau de urbanização	0=área urbana 1=área rural
RegOcup	Regime de Ocupação	1=proprietário, com crédito à habitação 2=proprietário, sem crédito à habitação 3=arrendatário (ou subarrend.) com renda a preços de mercado ou inferior 4=alojamento cedido gratuitamente ou a título de salário
Npagr	Número de pessoas no alojamento	0=1 ou 2 pessoas 1=3 ou + pessoas
País	País	0=Portugal 1=Espanha

Como sabemos, o número de classes (k) depende, entre outros, do número de observações de que dispomos ($n=10403$, em Portugal, $n=19435$, em Espanha). Calculado o número de classes seguiu-se a construção das mesmas, obedecendo às regras básicas de agrupamento dos valores em classes, referidas em Murteira *et al.* (1983).

Apesar de ser aconselhável, em termos genéricos, a amplitude das classes dever tomar valores constantes, pois facilita a comparação das frequências em cada uma das classes, nem sempre é possível cedermos a este princípio.

Tal como referido em Murteira *et al.* (1983), por serem dados relativos a rendimentos e despesas, pode não fazer sentido a amplitude de classes com valores constantes. Deste modo, foram consideradas amplitudes mais pequenas para as classes que representam os rendimentos e/ou despesas mais baixas, tendo sido adoptadas amplitudes mais elevadas para as últimas classes.

Na tabela 6 estão representados, para ano de referência 2005/2006, os rendimentos totais dos agregados familiares portugueses e o número de agregados em cada classe de rendimentos (IDEF 2005/2006, INE). Representámos o ponto médio de cada classe por x_i e n_i é o número de agregados em cada uma das classes (frequência absoluta).

Como sabemos a formação de classes é muito subjectiva e as classes que considerámos para as variáveis que considerámos neste trabalho podiam ter sido obviamente diferentes. Um pouco de bom senso, aliado a uma observação dos valores de vários percentis permitiram chegar às classes apresentadas.

Tabela 6: Classes da variável rendimento total, valores em € (Portugal)

Classes C_i	Ponto médio x_i	Frequência absoluta n_i
[0-6000)	3000	786
[6000-8000)	7000	809
[8000-10000)	9000	921
[10000-12000)	11000	898
[12000-14000)	13000	843
[14000-16000)	15000	741
[16000-19000)	17500	1150
[19000-22000)	20500	896
[22000-25000)	23500	656
[25000-30000)	27500	824
[30000-40000)	35000	933
[40000-50000)	42500	430
[50000-460690]	63470	516
	Total	10403

2.1.4 Curva de Lorenz, Índice de GINI

A partir daqui torna-se interessante definir e medir o grau de concentração dos rendimentos e das despesas de ambos os países.

Nos dados apresentados na Tabela 7 verificamos, numa primeira imagem, a distribuição dos rendimentos e a medida de concentração que é captada pela comparação das frequências acumuladas com nº de agregados correspondentes a cada classe, igualmente acumulada.

Tabela 7: Distribuição do rendimento total, valores apresentados em € (Portugal)

Classes C_i	$n_i \cdot x_i$	$p_i = n_i/n$	$n_i \cdot x_i$ acumulado	$Q_i = n_i x_i$ acumulado / total $n_i x_i$ acumulado	$P_i - Q_i$
[0-6000)	2358000	7,6%	2358000	1,1%	0,064
[6000-8000)	5663000	15,3%	8021000	3,8%	0,115
[8000-10000)	8289000	24,2%	16310000	7,8%	0,164
[10000-12000)	9878000	32,8%	26188000	12,6%	0,203
[12000-14000)	10959000	40,9%	37147000	17,8%	0,231
[14000-16000)	11115000	48,0%	48262000	23,1%	0,249
[16000-19000)	20125000	59,1%	68387000	32,8%	0,263
[19000-22000)	18368000	67,7%	86755000	41,6%	0,261
[22000-25000)	15416000	74,0%	102171000	49,0%	0,250
[25000-30000)	22660000	81,9%	124831000	59,9%	0,221
[30000-40000)	32655000	90,9%	157486000	75,5%	0,154
[40000-50000)	18275000	95,0%	175761000	84,3%	0,107
[50000-460690]	32750520	100,0%	208511520	100,0%	0,000
Total	208511520				

Através da análise dos dados obtidos verificamos, por exemplo, que aproximadamente 74% dos agregados familiares portugueses, os que têm rendimento total anual inferior a 25000 euros, auferem aproximadamente metade das remunerações recebidas pelo total dos agregados. Caso se verificasse igualdade na distribuição, as percentagens deveriam progredir simultaneamente. Repare-se, ainda, que os 18% dos agregados familiares portugueses que recebem mais de 25000 euros auferem 40% do total das remunerações recebidas.

À semelhança da variável referente a Portugal, foram calculados os valores relativos a Espanha, encontrando-se os valores na Tabela 8.

Tabela 8: Distribuição do rendimento total, valores das classes apresentados em € (Espanha)

Classes C_i	Ponto médio x_i	Frequência absoluta n_i	$n_i \cdot x_i$	$p_i = n_i/n$	$n_i \cdot x_i$ acumulado	$Q_i = n_i x_i$ acumulado / total $n_i x_i$ acumulado	$P_i - Q_i$
[0-6000)	3000	777	2330996	4,0%	2330996	0,5%	0,036
[6000-8000)	7000	693	4850997	7,6%	7181993	1,5%	0,061
[8000-10000)	9000	1075	9674995	13,1%	16856987	3,5%	0,096
[10000-12000)	11000	1360	14959993	20,1%	31816980	6,6%	0,135
[12000-14000)	13000	861	11192996	24,5%	43009976	8,9%	0,156
[14000-16000)	15000	1250	18749994	31,0%	61759970	12,8%	0,181
[16000-19000)	17500	2028	35489990	41,4%	97249960	20,2%	0,212
[19000-22000)	20500	1611	33025492	49,7%	130275452	27,1%	0,226
[22000-25000)	23500	1761	41383491	58,7%	171658943	35,7%	0,231
[25000-30000)	27500	2264	62259989	70,4%	233918932	48,6%	0,218
[30000-40000)	35000	2979	104264985	85,7%	338183917	70,2%	0,155
[40000-50000)	42500	1352	57459993	92,7%	395643910	82,2%	0,105
[50000-280000]	60249	1424	85794576	100,0%	481438486	100,0%	0,000
Total		19435	481438486				

E com base nos resultados, obtemos a seguinte representação da Figura 7 para a variável Rendimento Total, em Portugal e Espanha. Pode observar-se que as áreas de concentração são muito idênticas, sendo ligeiramente superior a área de Portugal, o que revela uma maior concentração. Contudo, o cálculo do IG para a variável em estudo, em Portugal e em Espanha, proporciona, sensivelmente o mesmo resultado.

$$IG_{PT} = \frac{6.38 - 4.09}{6.38} \approx 0.36$$

$$IG_{ES} = \frac{4.99 - 3.20}{4.99} \approx 0.36$$

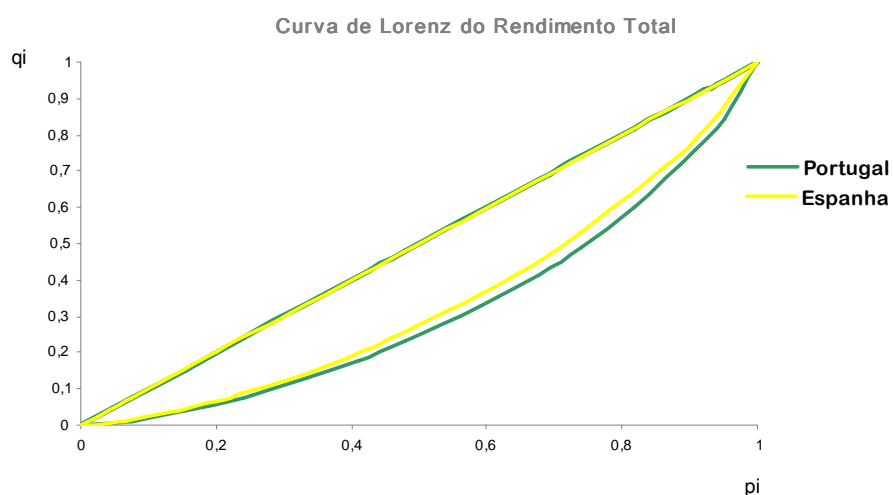


Figura 7: Curva de Lorenz, com os valores da variável Rendimento total, em Portugal e Espanha

Na Figura 8 representou-se a razão entre a mediana e 6 percentis (5, 10, 25, 75, 90 e 95) para Portugal e Espanha. Como podemos observar é na razão entre a mediana e os dois percentis mais elevados que existe uma maior diferença entre os dois países, sendo os valores mais elevados para Portugal. Portanto, é na distribuição dos rendimentos mais elevados (em especial, os 5% e 10% mais elevados) que se regista uma maior diferença entre os dois países.

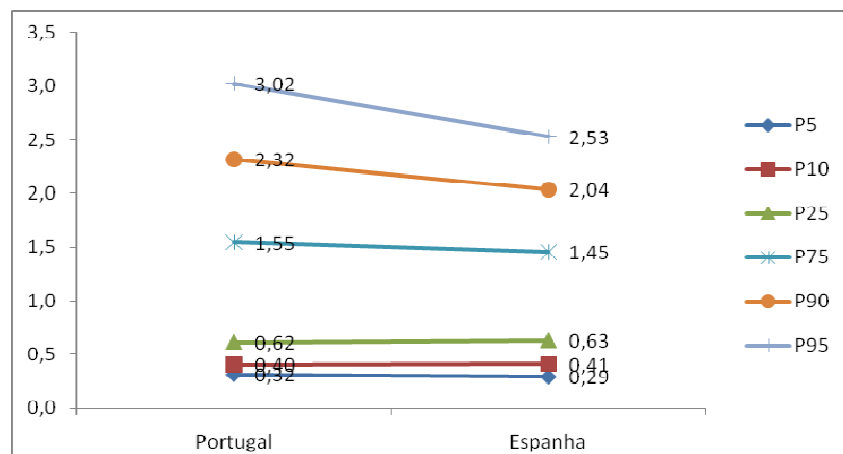


Figura 8: Comparação entre razões da mediana com os percentis 5%, 10%, 25%, 75%, 90% e 95% para o rendimento total em Portugal e Espanha

Rendimento monetário e rendimento não monetário

Para as variáveis rendimento monetário e não monetário apenas podemos apresentar resultados para Portugal (Tabelas 8 e 9), pois em Espanha não são divulgados os valores das mesmas.

O cálculo do IG para as variáveis em estudo proporciona, os seguintes resultados:

$$IG_{RM} = \frac{7.32 - 4.94}{7.32} \approx 0.33 \qquad IG_{RNM} = \frac{6.17 - 3.40}{6.17} \approx 0.47$$

Neste caso é de salientar uma maior concentração na distribuição dos rendimentos não monetários do que dos rendimentos monetários. Em particular, observe-se que os agregados familiares que auferem rendimentos não monetários superiores a 4000 euros detêm quase $\frac{3}{4}$ do rendimento não monetário ganho por todas as famílias.

Tabela 9: Distribuição do rendimento monetário, valores apresentados em € (Portugal)

Classes C_i	Ponto médio x_i	Frequência absoluta n_i	$n_i \cdot x_i$	$p_i = n_i/n$	$n_i \cdot x_i$ acumulado	$q_i = n_i x_i$ acumulado/total $n_i x_i$ acumulado	$p_i \cdot q_i$
[0-300)	150	1274	191100	12,2%	191100	0,5%	0,118
[300-1100)	700	525	367500	17,3%	558600	1,3%	0,160
[1100-2000)	1550	1039	1610450	27,3%	2169050	5,1%	0,222
[2000-2500)	2250	723	1626750	34,2%	3795800	9,0%	0,253
[2500-3000)	2750	439	1207250	38,5%	5003050	11,8%	0,266
[3000-3500)	3250	887	2882750	47,0%	7885800	18,7%	0,283
[3500-4000)	3750	898	3367500	55,6%	11253300	26,6%	0,290
[4000-4500)	4250	659	2800750	61,9%	14054050	33,2%	0,287
[4500-5000)	4750	773	3671750	69,4%	17725800	41,9%	0,274
[5000-6000)	5500	781	4295500	76,9%	22021300	52,1%	0,248
[6000-7000)	6500	908	5902000	85,6%	27923300	66,0%	0,196
[7000-8500)	7750	610	4727500	91,5%	32650800	77,2%	0,142
[8500-53366]	10855	887	9628385	100,0%	42279185	100,0%	0,000
Total		10403	42279185				

Tabela 10: Distribuição do rendimento não monetário, valores apresentados em € (Portugal)

Classes C_i	Ponto médio x_i	Frequência absoluta n_i	$n_i \cdot x_i$	$p_i = n_i/n$	$n_i \cdot x_i$ acumulado	$Q_i = n_i x_i$ acumulado/total $n_i x_i$ acumulado	$p_i - Q_i$
[0-300)	0	1274	0	12,2%	0	0,0%	0,122
[300-1100)	700	525	367500	17,3%	367500	0,9%	0,164
[1100-2000)	1550	1039	1610450	27,3%	1977950	4,7%	0,226
[2000-2500)	2250	723	1626750	34,2%	3604700	8,6%	0,257
[2500-3000)	2750	439	1207250	38,5%	4811950	11,4%	0,270
[3000-3500)	3250	887	2882750	47,0%	7694700	18,3%	0,287
[3500-4000)	3750	898	3367500	55,6%	11062200	26,3%	0,293
[4000-4500)	4250	659	2800750	61,9%	13862950	32,9%	0,290
[4500-5000)	4750	773	3671750	69,4%	17534700	41,7%	0,277
[5000-6000)	5500	781	4295500	76,9%	21830200	51,9%	0,250
[6000-7000)	6500	908	5902000	85,6%	27732200	65,9%	0,197
[7000-8500)	7750	610	4727500	91,5%	32459700	77,1%	0,144
[8500-53366]	10856	887	9628385	100,0%	42088085	100,0%	0,000
Total		10403	42088085				

Despesa

Relativamente à despesa total apresentam-se nas Tabelas 10 e 11 os valores obtidos após o agrupamento em classes.

Tabela 11: Distribuição da variável despesa total, valores apresentados em € (Portugal)

Classes C_i	Ponto médio x_i	Frequência absoluta n_i	$n_i \cdot x_i$	$p_i = n_i/n$	$n_i \cdot x_i$ acumulado	$Q_i = n_i x_i$ acumulado/total $n_i x_i$ acumulado	$p_i - Q_i$
[0-5000)	2500	981	2452495	9,4%	2452495	1,5%	0,079
[5000-7000)	6000	990	5939995	18,9%	8392490	5,1%	0,139
[7000-9000)	8000	1084	8671995	29,4%	17064485	10,3%	0,190
[9000-11000)	10000	1064	10639995	39,6%	27704479	16,8%	0,228
[11000-13000)	12000	1054	12647995	49,7%	40352474	24,5%	0,253
[13000-15000)	14000	884	12375996	58,2%	52728470	31,9%	0,263
[15000-17000)	16000	752	12031996	65,5%	64760466	39,2%	0,262
[17000-19000)	18000	627	11285997	71,5%	76046463	46,1%	0,254
[19000-21000)	20000	525	10499997	76,5%	86546460	52,4%	0,241
[21000-23000)	22000	406	8931998	80,4%	95478458	57,9%	0,226
[23000-26000)	24500	504	12347997	85,3%	107826456	65,3%	0,199
[26000-36000)	31000	871	27000996	93,6%	134827451	81,7%	0,120
[36000-196509]	45701	661	30208136	100,0%	165035588	100,0%	0,000
Total		10403	165035588				

Tabela 12: Distribuição da variável despesa total, valores apresentados em € (Espanha)

Classes C_i	Ponto médio x_i	Frequência absoluta n_i	$n_i \cdot x_i$	$p_i = n_i/n$	$n_i \cdot x_i$ acumulado	$Q_i = n_i x_i$ acumulado/total $n_i x_i$ acumulado	$p_i - Q_i$
[0-7000)	3500	582	2037000	3,0%	2037000	0,4%	0,026
[7000-10000)	8500	1073	9120500	8,5%	11157500	1,9%	0,066
[10000-13000)	11500	1353	15559500	15,5%	26717000	4,7%	0,108
[13000-16000)	14500	1521	22054500	23,3%	48771500	8,5%	0,148
[16000-19000)	17500	1509	26407500	31,1%	75179000	13,1%	0,179
[19000-22000)	20500	1549	31754500	39,0%	106933500	18,7%	0,204
[22000-25000)	23500	1519	35696500	46,9%	142630000	24,9%	0,219
[25000-28000)	26500	1406	37259000	54,1%	179889000	31,4%	0,227
[28000-31000)	29500	1283	37848500	60,7%	217737500	38,0%	0,226
[31000-34000)	32500	1092	35490000	66,3%	253227500	44,2%	0,221
[34000-39000)	36500	1541	56246500	74,2%	309474000	54,1%	0,202
[39000-44000)	41500	1231	51086500	80,6%	360560500	63,0%	0,176
[44000-419913]	56092	3776	211802835	100,0%	572363335	100,0%	0,000
Total		19435	572363335				

Neste caso, os valores do índice de Gini revelam uma maior concentração das despesas em Espanha do que em Portugal:

$$IG_{PT} = \frac{6.78 - 4.33}{6.78} \approx 0.36$$

$$IG_{ES} = \frac{5.03 - 3.03}{5.03} \approx 0.40$$

Na Figura 9 representou-se a razão entre a mediana e 6 percentis (5, 10, 25, 75, 90 e 95) para Portugal e Espanha. À semelhança do que verificámos para o rendimento é na razão entre a mediana e os dois percentis mais elevados que existe uma maior diferença entre os dois países, sendo os valores mais elevados para Portugal. Portanto, é na distribuição das despesas mais elevadas (em especial, os 5% mais elevados) que se regista uma maior diferença entre os dois países.

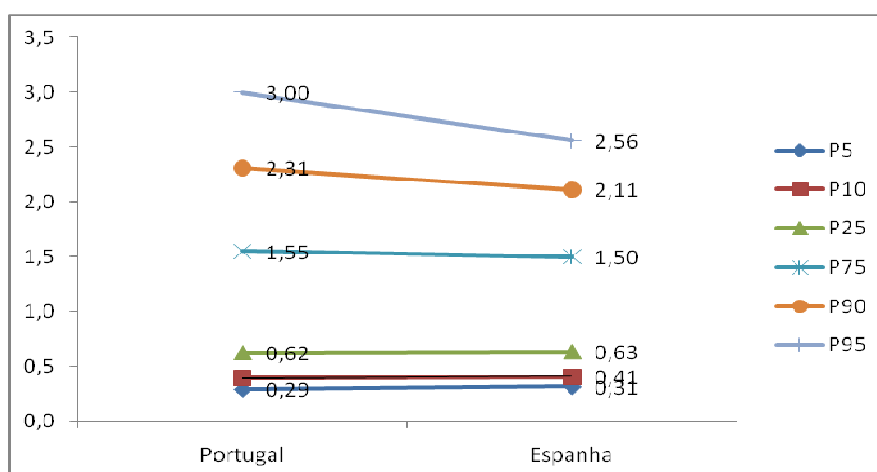


Figura 9: Comparação entre razões da mediana com os percentis 5%, 10%, 25%, 75%, 90% e 95% para a despesa total em Portugal e Espanha

Despesa Monetária

Relativamente à despesa monetária apresentam-se nas Tabelas 12 e 13 os valores obtidos após o agrupamento em classes, tendo-se obtido para o índice de Gini um maior valor para Espanha.

$$IG_{PT} = \frac{7.54 - 5.22}{7.54} \approx 0.31$$

$$IG_{ES} = \frac{5.71 - 3.45}{5.71} \approx 0.40$$

Tabela 13: Distribuição da variável despesa monetária, valores apresentados em € (Portugal)

Classes	Ponto médio	Frequência absoluta	$n_i \cdot x_i$	$p_i = n_i/n$	$n_i \cdot x_i$ acumulado	$q_i = n_i x_i$ acumulado/total	$p_i - q_i$
C_i	x_i	n_i				$n_i x_i$ acumulado	
[0-2000)	1000	561	561000	5,4%	561000	0,4%	0,049
[2000-3999,99)	3000	1218	3654000	17,1%	4215000	3,4%	0,137
[4000-5999,99)	5000	1342	6710000	30,0%	10925000	8,7%	0,213
[6000-7999,99)	7000	1341	9387000	42,9%	20312000	16,3%	0,266
[8000-9999,99)	9000	1199	10791000	54,4%	31103000	24,9%	0,295
[10000-12999,99)	11500	1334	15341000	67,2%	46444000	37,2%	0,301
[13000-15999,99)	14500	983	14253500	76,7%	60697500	48,6%	0,281
[16000-18999,99)	17500	661	11567500	83,0%	72265000	57,8%	0,252
[19000-23999,99)	21500	716	15394000	89,9%	87659000	70,2%	0,198
[24000-28999,99)	26500	375	9937500	93,5%	97596500	78,1%	0,154
[29000-33999,99)	31500	236	7434000	95,8%	105030500	84,1%	0,117
[34000-44999,99)	39500	255	10072500	98,3%	115103000	92,1%	0,061
[45000-150917]	54056	182	9838192	100,0%	124941192	100,0%	0,000
Total		10403	124941192				

Tabela 14: Classes da variável despesa monetária, valores das classes apresentados em € (Espanha)

Classes	Ponto médio	Frequência absoluta	$n_i \cdot x_i$	$p_i = n_i/n$	$n_i \cdot x_i$ acumulado	$q_i = n_i x_i$ acumulado/total	$p_i - q_i$
C_i	x_i	n_i				$n_i x_i$ acumulado	
[0-4000)	2000	722	1444000	3,7%	1444000	0,3%	0,034
[4000-7000)	5500	1338	7359000	10,6%	8803000	1,9%	0,087
[7000-10000)	8500	1559	13251500	18,6%	22064500	4,7%	0,140
[10000-13000)	11500	1695	19492500	27,3%	41547000	8,8%	0,186
[13000-16000)	14500	1698	24621000	36,1%	66168000	14,0%	0,221
[16000-19000)	17500	1649	28857500	44,6%	95025500	20,1%	0,245
[19000-22000)	20500	1547	31713500	52,5%	126739000	26,8%	0,257
[22000-26000)	24000	1890	45360000	62,2%	172099000	36,4%	0,258
[26000-30000)	28000	1516	42448000	70,0%	214547000	45,4%	0,247
[30000-34000)	32000	1198	38336000	76,2%	252883000	53,5%	0,227
[34000-39000)	36500	1192	43508000	82,3%	296391000	62,7%	0,197
[39000-44000)	41500	937	38885500	87,2%	335276500	70,9%	0,163
[44000-402584]	55149	2494	137541459	100,0%	472817959	100,0%	0,000
Total		19435	472817959				

Observe-se que para esta variável foi onde se obteve uma maior diferença na concentração entre Portugal e Espanha. Na Figura 10 representou-se a razão entre a mediana e 6 percentis (5, 10, 25, 75, 90 e 95) para Portugal e Espanha. À semelhança do que verificámos para o rendimento e para a despesa é na razão entre a mediana e os percentis mais elevados que existe uma maior diferença entre os dois países, sendo uma vez mais os valores mais elevados para Portugal.

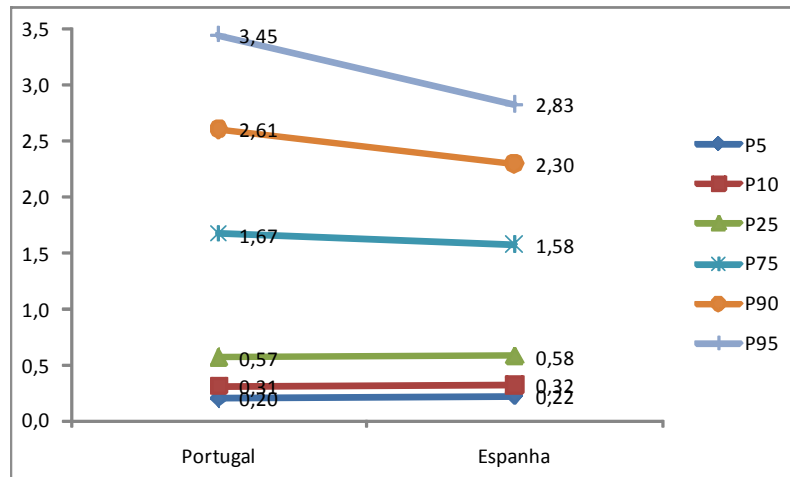


Figura 10: Comparação entre razões da mediana com os percentis 5%, 10%, 25%, 75%, 90% e 95% para a despesa monetária em Portugal e Espanha

Despesa Não Monetária

Relativamente à despesa monetária apresentam-se nas Tabelas 15 e 16 os valores obtidos após o agrupamento em classes.

Os valores obtidos para o índice de Gini foram mais uma vez desfavoráveis a Portugal.

$$IG_{PT} = \frac{6.17 - 3.44}{6.17} \approx 0.44$$

$$IG_{ES} = \frac{5.56 - 3.39}{5.56} \approx 0.39$$

Tabela 15: Distribuição da variável despesa não monetária, valores apresentados em € (Espanha)

Classes	Ponto médio x_i	Frequência absoluta n_i	$n_i \cdot x_i$	$p_i = n_i/n$	$n_i \cdot x_i$ acumulado	$Q_i = n_i x_i$ acumulado / total $n_i x_i$ acumulado	$p_i - Q_i$
[0-300)	150	1274	191100	12%	191100	0%	0,12
[300-1100)	700	525	367500	17%	558600	1%	0,16
[1100-2000)	1550	1039	1610450	27%	2169050	5%	0,22
[2000-2500)	2250	723	1626750	34%	3795800	9%	0,25
[2500-3000)	2750	439	1207250	38%	5003050	12%	0,27
[3000-3500)	3250	887	2882750	47%	7885800	19%	0,28
[3500-4000)	3750	898	3367500	56%	11253300	27%	0,29
[4000-4500)	4250	659	2800750	62%	14054050	33%	0,29
[4500-5000)	4750	773	3671750	69%	17725800	42%	0,27
[5000-6000)	5500	781	4295500	77%	22021300	52%	0,25
[6000-7000)	6500	908	5902000	86%	27923300	66%	0,20
[7000-8500)	7750	610	4727500	91%	32650800	77%	0,14
[8500-150917]	10855	887	9628385	100%	42279185	100%	0,00
Total		10403	42279185				

Tabela 16: Distribuição da variável despesa não monetária, valores apresentados em € (Espanha)

Classes	Ponto médio x_i	Frequência absoluta n_i	$n_i \cdot x_i$	$p_i = n_i/n$	$n_i \cdot x_i$ acumulado	$Q_i = n_i x_i$ acumulado / total $n_i x_i$ acumulado	$P_i - Q_i$
[0-20)	10	1799	17990	9%	17990	0%	0,09
[20-2500)	1260	628	791280	12%	809270	1%	0,12
[2500-3000)	2750	1053	2895750	18%	3705020	4%	0,14
[3000-3500)	3250	1507	4897750	26%	8602770	8%	0,17
[3500-4000)	3750	1679	6296250	34%	14899020	14%	0,20
[4000-4500)	4250	1678	7131500	43%	22030520	21%	0,22
[4500-5000)	4750	1544	7334000	51%	29364520	28%	0,23
[5000-5500)	5250	1494	7843500	59%	37208020	36%	0,23
[5500-6000)	5750	1402	8061500	66%	45269520	44%	0,22
[6000-6500)	6250	1137	7106250	72%	52375770	50%	0,21
[6500-7500)	7000	1564	10948000	80%	63323770	61%	0,19
[7500-9000)	8250	1406	11599500	87%	74923270	72%	0,15
[9000-65258]	11408	2544	29021675	100%	103944945	100%	0,00
Total		19435	103944945				

Contudo, ao representarmos a razão da mediana com os 6 percentis (Figura 11) podemos observar que para 10% e 25% se verifica um aumento de Portugal para Espanha (maior desigualdade em Espanha para despesas não monetárias mais pequenas) enquanto para os percentis maiores se verifica uma variação inversa, sendo que para despesas não monetárias muito maiores há uma maior desigualdade em Portugal.

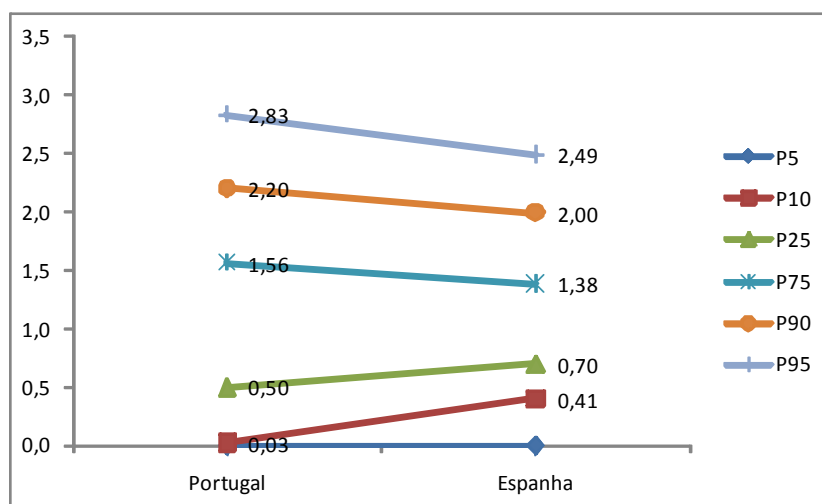


Figura 11: Comparação entre razões da mediana com os percentis 5%, 10%, 25%, 75%, 90% e 95% para a despesa não monetária em Portugal e Espanha

Sendo que, nos resultados apresentados até ao momento foram assumidos como representativos os pontos médios das classes, não poderíamos prosseguir a análise sem antes verificar se seriam obtidos resultados semelhantes com os valores reais acumulados em cada classe, medir o erro relativo associado e calcular os valores do Índice de Gini.

Os valores do Índice de Gini, obtidos para as variáveis em estudo podem ser analisados na Tabela 18. Quando comparados com os valores apresentados anteriormente, podemos destacar que se obtiveram valores muito semelhantes nas duas abordagens. Para Portugal, o valor do índice de concentração manteve-se inalterável para as variáveis despesa total e despesa monetária, tendo-se verificado o maior aumento (0.03) para a despesa não monetária.

Tabela 17: Comparação dos resultados do Índice de Gini obtidos através dos valores reais e em classes, para Portugal e Espanha

	Valores Reais						Valores Classes					
	pi	qi	IG _{PT}	pi	qi	IG _{ES}	pi	qi	IG _{PT}	pi	qi	IG _{ES}
Rendimento total	6,38	4,09	0,36	4,99	3,20	0,36	6,37	4,04	0,37	4,99	3,13	0,37
Rendimento monetário	7,32	4,84	0,34	-	-		7,32	4,94	0,33	-	-	
Rendimento não monetário	6,17	3,29	0,47	-	-		6,17	3,40	0,45	-	-	
Despesa total	6,78	4,31	0,36	5,03	2,93	0,42	6,78	4,33	0,36	5,03	3,03	0,40
Despesa monetária	7,54	5,21	0,31	5,71	3,36	0,41	7,54	5,22	0,31	5,71	3,45	0,40
Despesa não monetária	6,17	3,29	0,47	5,66	3,33	0,40	6,17	3,44	0,44	5,66	3,39	0,39

Na Tabela 18 podemos verificar que os erros de agrupamento cometidos são muito pequenos no caso da média e são inferiores a 25% no caso do desvio padrão para as variáveis deste estudo de concentração.

Tabela 18: Comparação dos resultados do valor da média e do desvio padrão obtidos através dos valores não agrupados e agrupados em classes para Portugal

Variável	Média			Desvio Padrão		
	Agrupados	Não Agrupados	Erro Relativo	Agrupados	Não Agrupados	Erro Relativo
Despesa Total	16184	15864	-2%	12247	10947	-12%
Despesa Monetária	12027	12010	0%	10477	10051	-4%
Despesa Não Monetária	4157	4064	-2%	2659	2981	11%
Rendimento Total	20615	20043	-3%	17536	14114	-24%
Rendimento Monetário	16459	16186	-2%	15977	12992	-23%
Despesa Não Monetária	4157	4046	-3%	3659	2981	-23%

2.1.5 Análise descritiva de outras variáveis

Na modelação do capítulo seguinte, além dos rendimentos, consideramos outras variáveis explicativas como o grau de urbanização, alojamento, o número de pessoas do agregado familiar e índice de conforto total. Neste ponto fazemos uma análise exploratória de dados a cada uma delas.

Grau de Urbanização

A variável Grau de urbanização tem três categorias e representam a desagregação territorial. Em 2005/2006, cerca de 63% dos agregados familiares inquiridos residiam em áreas predominantemente urbanas. A proporção de agregados familiares a residir em áreas mediamente urbana e predominantemente rurais é semelhante, contudo apresentam valores inferiores a 20% (Figura 12).

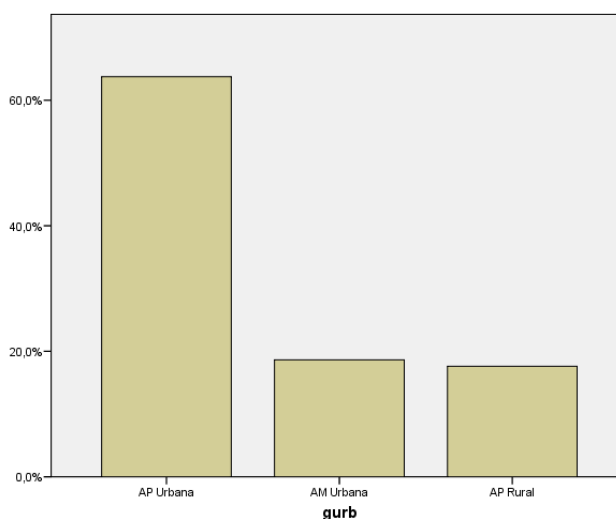


Figura 12: Distribuição da variável grau de urbanização por categorias

Alojamento

Relativamente ao alojamento, verificamos que a maior parte dos agregados familiares residiam em alojamentos dos quais eram proprietários, sendo que 57% do total de inquiridos não possuía crédito à habitação, contudo 20% dos agregados familiares recorreram ao crédito para financiamento da habitação. Abaixo dos 10% estavam representadas as categorias de arrendatários (com rendas a preço de mercado ou inferiores) inquiridos cujo alojamento era cedido gratuitamente ou a título de salário. Na Figura 13 podemos observar o gráfico de barras com a distribuição por categorias.

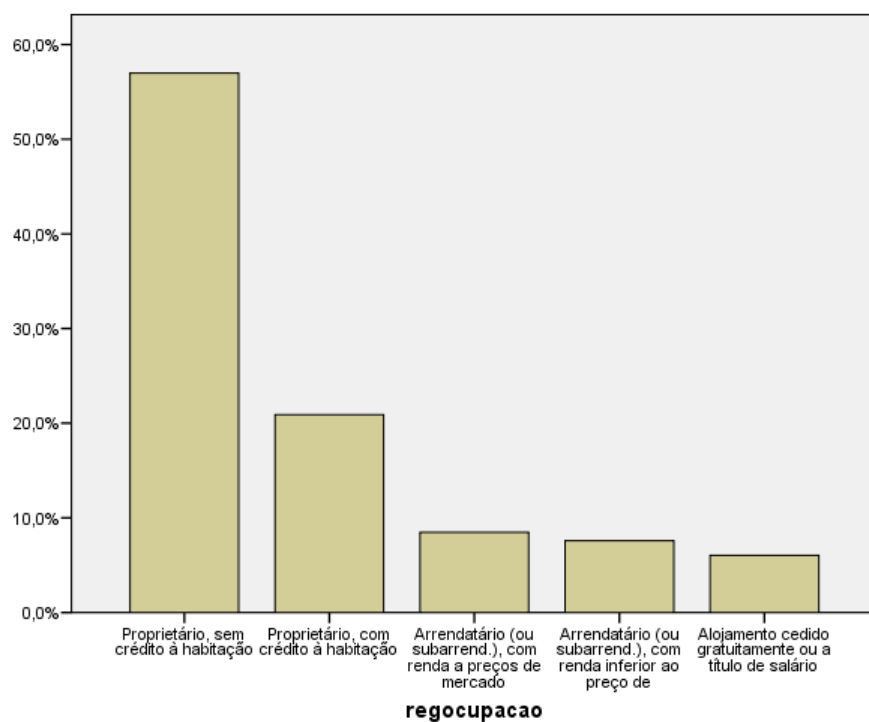


Figura 13: Distribuição da variável alojamento por categorias

Número de Pessoas do Agregado Familiar

A maioria dos agregados familiares inquiridos (32%), eram constituídos por 2 pessoas. Aproximadamente 8% dos agregados familiares eram constituídos por 5 ou mais pessoas. Na Tabela 19 representam-se as frequências observadas por composição do agregado familiar.

Tabela 19: Tabela de frequências para o número de pessoas por agregado familiar

		Nº	%	% válida	% acumulada
Valid	1 pessoas	1831	17,6	17,6	17,6
	2 pessoas	3399	32,7	32,7	50,3
	3 pessoas	2404	23,1	23,1	73,4
	4 pessoas	1870	18,0	18,0	91,4
	5 ou + pessoas	899	8,6	8,6	100,0
	Total	10403	100,0	100,0	

Índice de Conforto Total

Através dos dados utilizados na análise tornava-se possível identificar o nº de agregados familiares que dispunha dos bens ou serviços descritos no capítulo 2. Todavia, dado que a publicação dos dados do inquérito (INE, 2008) disponibilizava algumas informações, por tipos de bens e/ou equipamentos, optou-se por utilizar a estrutura de dados de que dispúnhamos para obter um índice, cujos resultados se encontram no histograma da Figura 14. Este processo possibilitou a utilização de um conjunto de informações numa só variável.

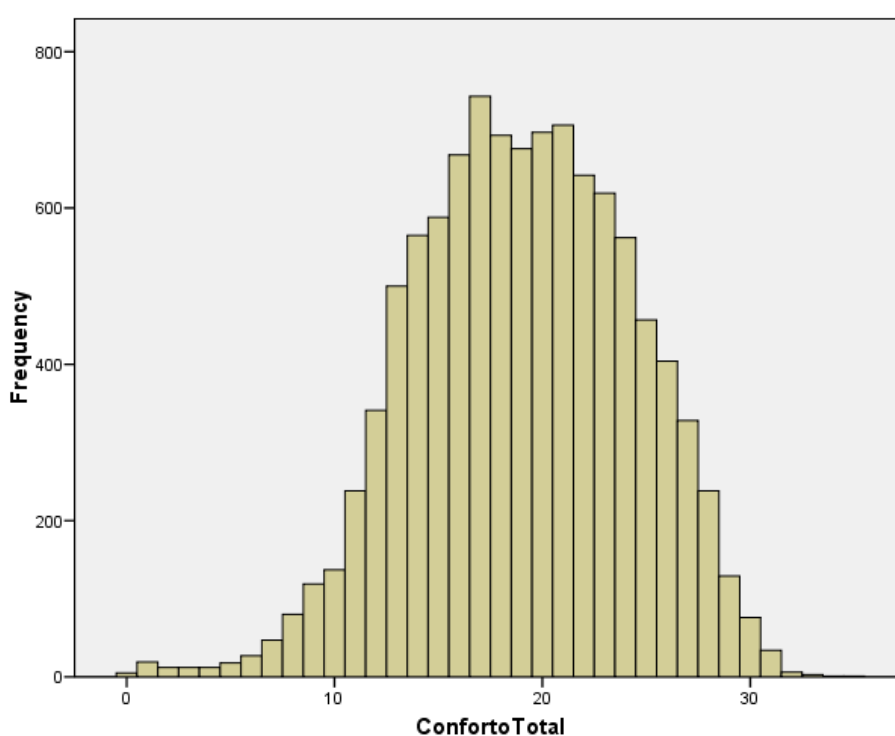


Figura 14: Histograma para o índice de conforto total.

Saliente-se que nenhum dos agregados familiares inquiridos dispunha de todos os bens ou serviços incluídos no inquérito. Apenas um agregado familiar inquirido dispunha de 35 dos 36 bens ou equipamentos incluídos no questionário, contudo foram inquiridos 5 agregados familiares que afirmavam não dispor de nenhum dos equipamentos ou bens. Metade dos agregados possui o número médio (19) de bens ou equipamentos, registando-se um desvio padrão igual a 5,25 e sendo a distribuição assimétrica negativa e leptocúrtica.

Média	19,02
Mediana	19,00
Desvio padrão	5,25
Assimetria	-0,18
Erro-padrão (Assimetria)	0,02
Curtose	-0,17
Erro-padrão (Curtose)	0,05

Capítulo III

3. Modelação Estatística da Despesa dos Agregados

Neste capítulo procuramos averiguar a influência das variáveis rendimento, regime de ocupação, número de pessoas do agregado familiar, grau de urbanização e conforto na despesa dos agregados, a qual foi dicotomizada usando a mediana como ponto de corte.

3.1 Modelo de Regressão Logística para Portugal

Consideramos a variável resposta como a despesa categorizada em que o evento de interesse é o de determinado agregado familiar ter um rendimento acima da mediana. Na Tabela 20 podemos observar o ajustamento do modelo com cada variável individualmente, onde podemos constatar a significância de todas as variáveis.

Tabela 20: Resultados do ajustamento de modelos com cada variável individualmente

Efeito	β	Desvio padrão	Wald	valor de p	Exp(β)	95% IC Exp(β)	
						Limite inferior	Limite superior
RendTot(1)	2,76	0,05	3182	<0,001	15,80	14,36	17,39
RendMonet(1)	2,30	0,05	2506	<0,001	9,93	9,08	10,86
RendNaoMonet(1)	1,95	0,04	1962	<0,001	7,02	6,44	7,65
NPAgr(1)	1,69	0,04	1560	<0,001	5,41	4,98	5,89
RegOcup			748	<0,001			
RegOcup(1)	-1,31	0,06	536	<0,001	0,27	0,24	0,30
RegOcup(2)	-1,81	0,07	635	<0,001	0,16	0,14	0,19
RegOcup(3)	-1,54	0,10	261	<0,001	0,21	0,18	0,26
GUrb(1)	-0,90	0,05	270	<0,001	0,41	0,37	0,45
GUrb_inicial			326	<0,001			
GUrb_inicial(1)	-0,39	0,05	56	<0,001	0,68	0,61	0,75
GUrb_inicial(2)	-0,98	0,06	311	<0,001	0,37	0,34	0,42
Conftotal	0,31	0,01	2425	<0,001	1,36	1,35	1,38

A partir dos resultados obtidos com os vários modelos podemos tecer as seguintes considerações:

- Quando os agregados familiares auferem um total de rendimentos acima da mediana, a possibilidade de terem despesas acima da mediana aumenta 15 vezes face aos agregados que auferem rendimentos abaixo da mediana.

- Quando os agregados familiares auferem rendimentos monetários acima da mediana, a possibilidade de terem despesas acima da mediana aumenta 10 vezes face aos agregados que auferem rendimentos abaixo da mediana.
- Quando os agregados familiares auferem rendimentos não monetários acima da mediana, a possibilidade de terem despesas acima da mediana aumenta 7 vezes face aos agregados que auferem rendimentos abaixo da mediana.
- Quando o número de indivíduos do agregado é composto por 3 ou mais pessoas a possibilidade de terem despesas acima da mediana aumenta 5 vezes face aos agregados que são compostos por 1 ou duas pessoas.
- Quando os residentes são proprietários do alojamento possuem crédito à habitação a possibilidade de terem despesas acima da mediana diminui cerca de 2% face aos residentes que são proprietários sem crédito à habitação.
- Quando os alojamentos são constituídos por arrendatários, a possibilidade de terem despesas acima da mediana diminui face aos residentes que são proprietários sem crédito à habitação.
- Quando os residentes ocupam alojamentos cedidos gratuitamente ou a título de salário, a possibilidade de terem despesas acima da mediana diminui face aos residentes que são proprietários sem crédito à habitação.
- Quando os agregados familiares residem em áreas rurais, a possibilidade de terem despesas acima da mediana é menor face aos agregados familiares residentes em áreas urbanas.
- Quando os agregados familiares residem em áreas mediantemente urbanas, a possibilidade de terem despesas acima da mediana é menor face aos agregados familiares residentes em áreas predominantemente urbanas.
- Quando os agregados familiares residem em áreas predominantemente rurais, a possibilidade de terem despesas acima da mediana é menor face aos agregados familiares residentes em áreas predominantemente urbanas.
- Quando o número de bens ou serviços aumenta uma unidade a possibilidade de terem uma despesa acima da mediana aumenta 35%.

Começámos por ajustar um modelo com todas as variáveis exceptuando, naturalmente, o caso do rendimento, já que a soma dos rendimentos monetário e não monetário é o rendimento total. Assim, tomámos as duas abordagens estando os resultados apresentados na Tabela 21, da qual podemos salientar o seguinte:

- (i) O efeito da covariável *regocup*, não se revela estatisticamente significativo nas categorias 2 e 4 relativamente à categoria base.
- (ii) Com a entrada dos rendimentos desagregados no modelo em vez do rendimento total, também a 2ª categoria da variável *gurb* deixa de ser significativa.

Tabela 21: Modelos de regressão logística ajustados para Portugal e respectivas tabelas de resultados

Variável	Coefficientes	Desvio padrão	z	Pr(> z)
Constante	-4,99	0,14	-35,47	< 2e-16
rendtot	1,71	0,06	30,35	< 2e-16
npagr1	0,78	0,06	14,04	< 2e-16
regocup2	-0,05	0,07	-0,64	0,5196
regocup3	-0,39	0,08	-5,17	2.39e-07
regocup4	-0,15	0,11	-1,34	0,18
gurb2	-0,13	0,07	-1,83	0,07
gurb3	-0,45	0,08	-5,88	4.01e-09
conftotal	0,20	0,01	28,32	< 2e-16

Variável	Coefficientes	Desvio padrão	z	Pr(> z)
Constante	-5,61	0,15	-37,75	< 2e-16
rmonet1	1,25	0,06	21,63	< 2e-16
rnaomonet1	1,63	0,06	26,70	< 2e-16
npagr1	0,90	0,06	15,81	< 2e-16
regocup2	-0,09	0,07	-1,15	0,25
regocup3	0,15	0,08	1,87	0,06
regocup4	-0,13	0,12	-1,12	0,26
gurb2	-0,06	0,07	-0,86	0,39
gurb3	-0,32	0,08	-4,14	3.56e-05
conftotal	0,20	0,01	26,90	< 2e-16

Nesta fase estão identificadas as covariáveis significativas em ambos os modelos, mas dadas as diferenças é importante decidir com qual prosseguimos a análise. Como o modelo com a desagregação do rendimento tem menor AIC decidimos prosseguir para a fase de averiguação de possíveis interações usando esse modelo. Contudo, como uma das categorias da variável *gurb* não é significativa decidimos recodificação da mesma:

Variável	Descrição	Codificação
Gurb	Grau de urbanização	1=área predominantemente urbana 2=área mediamente urbana 3=área predominantemente rural

Variável	Codificação
Gurb2	0=área urbana 1=área rural

Nesta fase, este procedimento foi aplicado à covariável *grau de urbanização*:

De seguida foi seleccionada a covariável que apresentou o menor valor de p para se dar início à pesquisa por inclusão de possíveis interacções significativas e posteriormente à verificação do seu contributo. Identificámos alguns modelos possíveis e comparámo-los com o obtido pelo procedimento “*step*” do R, o qual selecciona as variáveis de acordo com o AIC. Este método consiste em identificar todas as covariáveis e as respectivas interacções, independentemente dos seus efeitos serem ou não significativos, sendo depois ajustados todos os possíveis modelos e para cada um calculado o valor do AIC. Aquele que apresentar um menor valor de AIC é o seleccionado.

A análise deve ser feita na tentativa de encontrar o modelo que apresente os coeficientes com o melhor ajustamento possível, mas que seja parcimonioso. Na Tabela 22 apresentamos um modelo possível, mas que tem imensos parâmetros e consequentemente resulta numa maior dificuldade de interpretação. Por outro lado, quando comparado com o modelo apresentado na Tabela 23 tem uma capacidade explicativa muito próxima com valores de AIC e *Deviance* próximos.

Tabela 22: Modelo de regressão logística ajustado para Portugal

Variável	Coefficientes	Desvio padrão	z	Pr(> z)
Constante	-6,50	0,25	-26,49	< 2e-16
rmonet1	1,32	0,10	13,47	< 2e-16
rmaomonet1	1,60	0,08	21,14	< 2e-16
npagr1	2,18	0,29	7,41	1.24e-13
regocup2	0,77	0,44	1,75	0,08
regocup3	0,67	0,39	1,69	0,09
regocup4	0,34	0,56	0,61	0,54
conftotal	0,25	0,01	19,34	< 2e-16
npagr1:conftotal	-0,05	0,01	-3,67	0,00
regocup2:conftotal	-0,05	0,02	-2,32	0,02
regocup3:conftotal	-0,05	0,02	-2,32	0,02
regocup4:conftotal	-0,03	0,03	-1,19	0,23
npagr0:gurb2	-0,08	0,12	-0,65	0,51
npagr1:gurb2	-0,13	0,11	-1,12	0,26
npagr0:gurb3	-0,34	0,12	-2,75	0,01
npagr1:gurb3	-0,57	0,13	-4,52	0,00
rmonet1:npagr1	-0,37	0,12	-3,12	0,00
regocup2:gurb2	-0,12	0,19	-0,63	0,53
regocup3:gurb2	0,28	0,22	1,26	0,21
regocup4:gurb2	0,33	0,28	1,16	0,25
regocup2:gurb3	0,18	0,21	0,84	0,40
regocup3:gurb3	0,59	0,30	1,94	0,05
regocup4:gurb3	0,97	0,31	3,11	0,00
rmonet1:regocup2	0,20	0,15	1,30	0,19
rmonet1:regocup3	0,38	0,15	2,47	0,01
rmonet1:regocup4	-0,09	0,24	-0,38	0,71
rmaomonet1:regocup2	-0,03	0,15	-0,18	0,86
rmaomonet1:regocup3	0,83	0,37	2,27	0,02
rmaomonet1:regocup4	-0,10	0,23	-0,43	0,67

Aplicámos o teste da razão de verosimilhanças (os dois modelos são encaixados) que deu significativo a favor do modelo apresentado na Tabela 22. Contudo, optámos por escolher o modelo apresentado na Tabela 23 pela sua muito maior simplicidade.

Tabela 23: MRL final, para Portugal

Efeito	β	Desvio padrão	Wald	valor de p	Exp(β)	95% IC Exp(β)	
						Limite inferior	Limite superior
RendMonet(1)	1,43	0,09	269,70	0,00	4,16	3,51	4,94
RendNaoMonet(1)	1,62	0,06	709,02	0,00	5,06	4,49	5,71
Conftotal	0,23	0,01	412,41	0,00	1,26	1,23	1,28
GUrb(1)	-0,30	0,08	15,58	0,00	0,74	0,64	0,86
NPAgr(1)	2,17	0,28	60,71	0,00	8,73	5,06	15,06
RegOcup			6,30	0,10			
RegOcup(1)	0,07	0,07	0,98	0,32	1,08	0,93	1,25
RegOcup(2)	0,20	0,10	4,51	0,03	1,23	1,02	1,48
RegOcup(3)	-0,06	0,13	0,25	0,62	0,94	0,73	1,20
Conftotal by NPAgr(1)	-0,06	0,01	15,06	0,00	0,95	0,92	0,97
NPAgr(1) by RendMonet(1)	-0,36	0,12	9,61	0,00	0,70	0,56	0,88

3.1.1 Bondade do Ajustamento e Diagnóstico

A cováriavel *conftotal* foi modelada como variável contínua e, tal como verificámos, foi incluída no modelo final. Quando o modelo contém uma cováriavel contínua, deve ser assegurada a sua linearização com a função logit.

Usualmente faz-se uma representação gráfica dos pontos médios dos intervalos obtidos pela categorização da variável contínua usando os quartis como ponto de corte contra os coeficientes de cada categoria (a de referência naturalmente com o valor nulo) do modelo ajustado, substituindo a variável contínua pela variável contínua categorizada. Naturalmente, espera-se que desta representação gráfica resulte aproximadamente uma recta, e é o que acontece para este modelo, como ilustra a Figura 15.

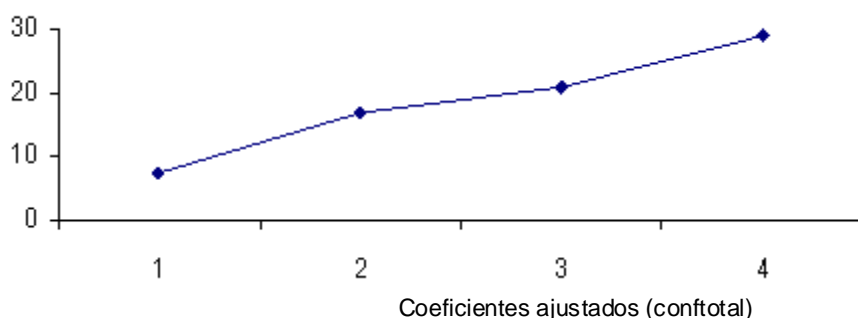


Figura 15: Representação dos coeficientes ajustados pelo modelo para as categorias da cováriavel contínua *conftotal* categorizada

Verificado este pressuposto recorremos ao teste de Hosmer e Lemeshow para testar bondade de ajustamento do modelo, obtendo-se um valor da estatística qui-quadrado igual a 2.76, o que para 8 graus de liberdade permite aceitar a hipótese de bom ajustamento, uma vez que tem associado um valor p igual a 0,95.

Através da representação gráfica dos resíduos podemos procurar a existência de observações individuais que possam, de alguma forma, ter influência sobre a estimação dos parâmetros ou de observações inconsistentes (*outliers*).

Do gráfico da Figura 16, que representa os resíduos de Pearson, podemos destacar duas observações que se afastam das restantes.

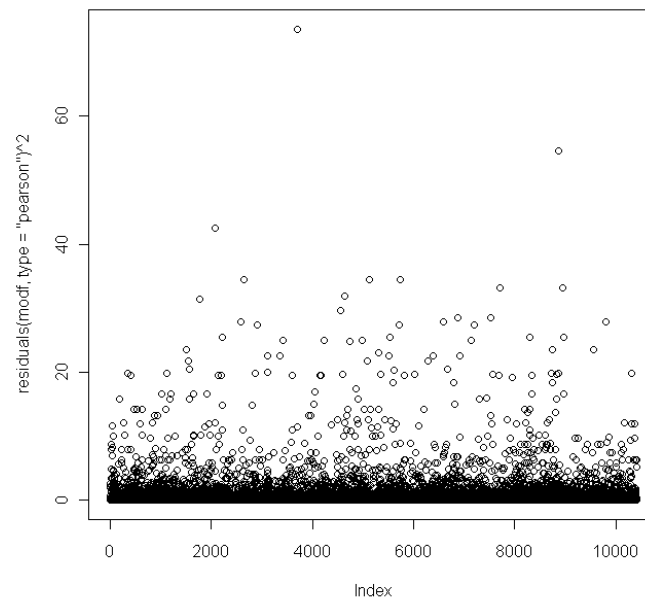


Figura 16: Resíduos de Pearson

Graficamente poderemos avaliar o modelo através dos resíduos *deviance*, os quais estão representados na Figura 17, não se registando nada de anormal na sua representação e estando todos no intervalo entre -3 e 3.

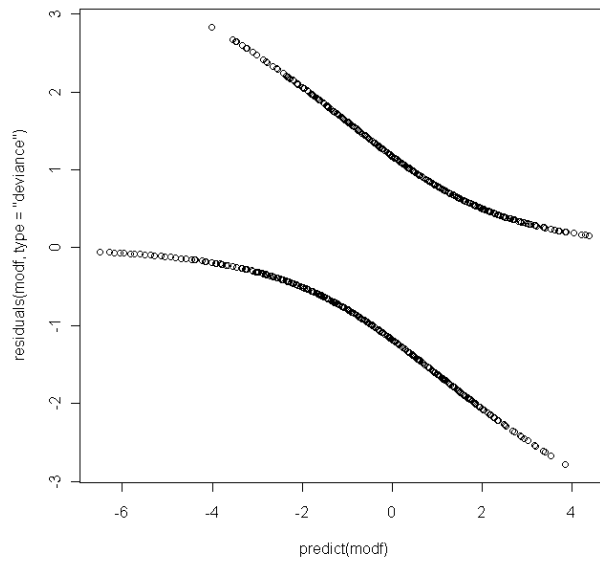


Figura 17: Resíduos *Deviance*

A influência de cada observação na estimação de cada coeficiente do modelo pode ser estimada pelos resíduos DFbetas (Figura 18) correspondendo cada figura à representação dos valores de cada covariável com os resíduos correspondentes. Da observação da Figura 4 não temos nada de especial a salientar.

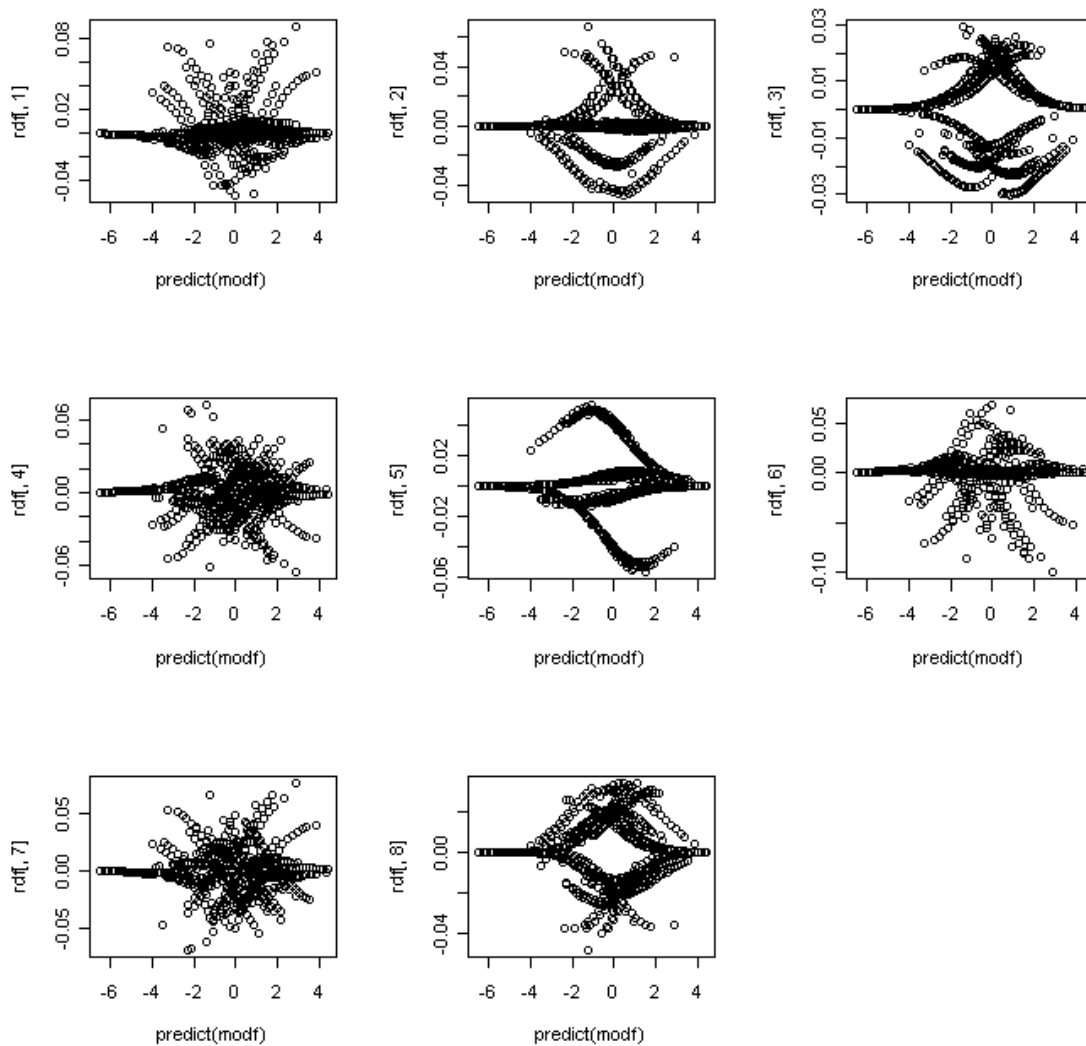


Figura 18: Resíduos DfBetas

Procedemos de seguida à representação da Distância de Cook. Como podemos verificar na Figura 19 não existem pontos acima de 0.5 que poderia denunciar a existência de observações influentes. Contudo, existem observações que se destacam devido ao seu afastamento e que, por esse motivo devem ser investigadas, pois a sua presença pode ter consequências nas conclusões a retirar através do modelo. Neste caso as observações que mais se afastam são as 4372, 7697 e 6224.

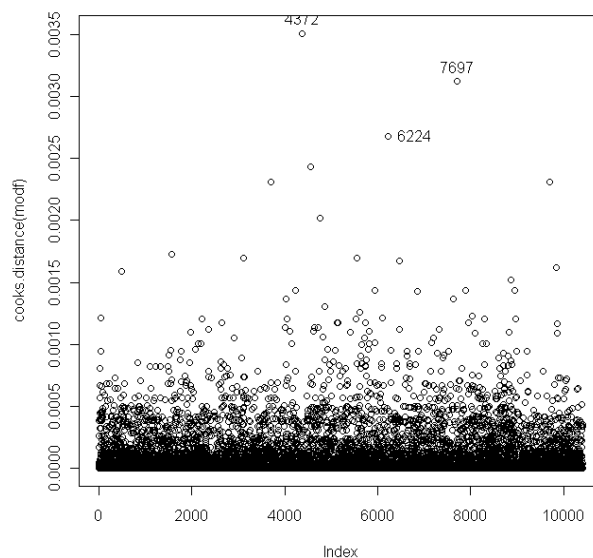


Figura 19: Distância de Cook

É necessário identificar o efeito destas observações nos parâmetros estimados, de modo a verificar o quão influente essa observação é. Para tal ajustou-se o modelo com e sem estas observações tendo-se obtido apenas pequenas variações nos valores dos parâmetros e da *deviance* do modelo, pelo que optámos por as deixar ficar no modelo.

Conjuntamente com as representações gráficas para verificação da adequação do modelo devem ser analisadas as medidas de sensibilidade e especificidade.

Neste modelo, o valor de corte que maximiza a sensibilidade e especificidade é 0.5, mas no entanto são apresentados, na Tabela 24 alguns valores obtidos de sensibilidade e especificidade com diferentes pontos de corte.

Tabela 24: Valores de sensibilidade e especificidade obtidos, com diferentes pontos de corte para o modelo de Portugal

Quando o ponto de corte é	0,4	0,45	0,5	0,52	0,55	0,6
sensibilidade	88%	84%	82%	81%	79%	75%
especificidade	73%	77%	79%	80%	82%	85%

O objectivo deste teste, no âmbito do estudo, é evitar predizer uma não ocorrência incorrectamente (falso negativo), desta forma, é o ponto de corte de 0.4 que

proporciona uma percentagem mais elevada de sensibilidade (i.e. percentagem de ocorrências correctamente preditas). Deste modo, a probabilidade de existirem agregados familiares com despesas acima da mediana, correctamente classificados é 88%.

Na Figura 20 representámos a Curva ROC obtida. Este modelo tem uma capacidade classificatória quase excelente, pois a área estimada abaixo da curva foi igual a 0.889, estando, para um nível de confiança de 95%, entre 0.883 e 0.895.

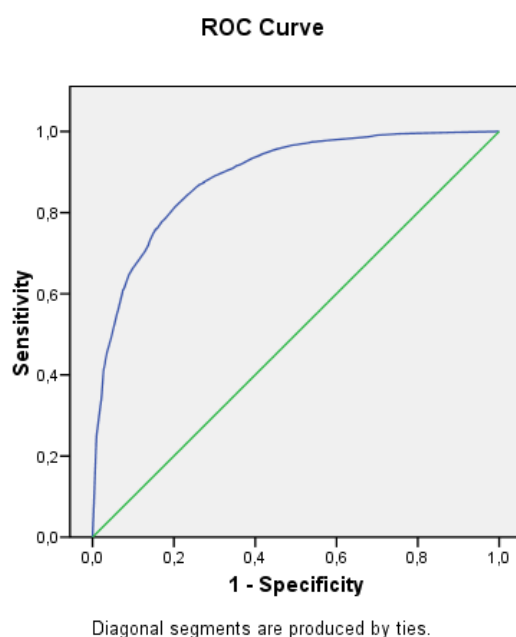


Figura 20: Curva ROC para o modelo logístico das despesas dos agregados portugueses.

3.1.2 Interpretação dos Coeficientes

A partir dos coeficientes do modelo obtido (Tabela 22) é possível retirar as seguintes conclusões (sempre admitindo fixas as restantes covariáveis):

- Quando os agregados familiares auferem um total de rendimentos monetário acima da mediana, a possibilidade de terem despesas acima da mediana aumenta 4 vezes face aos agregados que auferem rendimento monetário abaixo da mediana.
- Os agregados que residem em áreas rurais têm 25% menos de possibilidades de ter despesas acima da mediana.

- Quando os residentes são proprietários do alojamento e possuem crédito à habitação a possibilidade de terem despesas acima da mediana aumenta cerca de 8% face aos residentes que são proprietários sem crédito à habitação.
- Quando os alojamentos são constituídos por arrendatários, a possibilidade de terem despesas acima da mediana aumenta 23% face aos residentes que são proprietários sem crédito à habitação.
- Quando os residentes ocupam alojamentos cedidos gratuitamente ou a título de salário, a possibilidade de terem despesas acima da mediana diminui 6% face aos residentes que são proprietários sem crédito à habitação.

Como as interações entre o número de pessoas do agregado familiar e o conforto total e o rendimento monetário são significativas então a interpretação é feita de uma outra forma, uma vez que devemos interpretar uma variável tendo fixos os níveis da outra. Assim, por exemplo, deixamos os cálculos efectuados para o caso do Odds Ratio da interacção entre as variáveis NPAgr e RendMonet. Fixando o Rendimento Monetário na classe acima da mediana (RendMonet=1) tem-se

$$\begin{aligned} & \text{logit}(\text{RendMonet} = 1, \text{NPAgr} = 1) - \text{logit}(\text{RendMonet}=1, \text{NPAgr}=0) = \\ & = (\beta_0 + \beta_1 + \beta_5 + \beta_{11}) - (\beta_0 + \beta_1) = \\ & = \beta_5 + \beta_{11} = 2.17 - 0.36 = 1.81 \end{aligned}$$

Logo, $OR = \exp(1,81) = 6,1$. Após o cálculo da variância, que se faz recorrendo à matriz de variâncias covariâncias, o intervalo de confiança é dado por:

$$\exp(\hat{\beta}_5 + \hat{\beta}_{11}) \pm 1.96\sigma_{\hat{\beta}_5 + \hat{\beta}_{11}} = (5.3; 7.2) .$$

Assim:

- Para os agregados com rendimento monetário acima da mediana, caso o número de pessoas do agregado seja 3 ou mais, as possibilidades de terem uma despesa acima da mediana relativamente a agregados com uma ou duas pessoas aumenta 6 vezes (IC_{95%}: 5,3; 7,2) enquanto para os agregados com rendimento abaixo da mediana a mesma razão de possibilidades aumenta 9 vezes.
- Para um agregado com uma ou duas pessoas, por cada bem ou serviço a mais a possibilidade de terem uma despesa acima da mediana aumenta 26%, enquanto nos agregados com 3 ou mais pessoas essa possibilidade aumenta 7%.

3.2 Modelo de Regressão Logística para Espanha

Consideramos agora os valores da despesa categorizada em que o evento de interesse é determinado por os agregado familiar espanhol ter um rendimento acima da mediana. Na Tabela 25 podemos observar o ajustamento do modelo com cada variável individualmente, onde podemos constatar a significância de todas as variáveis. Recorde-se que em Espanha apenas é disponibilizado o valor do Rendimento Total e não da sua desagregação em Monetário e Não Monetário.

Tabela 25: Resultados do ajustamento de modelos com cada variável individualmente (Espanha)

Efeito	β	Desvio padrão	Wald	valor de p	Exp(β)	95% IC Exp(β)	
						Limite inferior	Limite superior
RendTot(1)	2,18	0,03	4352	<0,001	8,86	8,30	9,45
NPAgr(1)	1,62	0,03	2652	<0,001	5,07	4,77	5,40
Gurb			232	<0,001			
Gurb(1)	-0,09	0,04	6	0,013	0,91	0,85	0,98
Gurb(2)	-0,50	0,03	222	<0,001	0,61	0,57	0,65
RegOcup			935	<0,001			
RegOcup(1)	0,81	0,03	556	<0,001	2,25	2,10	2,41
RegOcup(2)	-0,69	0,05	174	<0,001	0,50	0,45	0,56
RegOcup(3)	-0,25	0,07	14	<0,001	0,78	0,68	0,89

A partir dos resultados contidos podemos tecer as seguintes considerações:

- Quando os agregados familiares auferem um total de rendimentos acima da mediana, a possibilidade de terem despesas acima da mediana aumenta cerca de 9 vezes face aos agregados que auferem rendimentos abaixo da mediana.
- Quando o número de indivíduos do agregado é composto por 3 ou mais pessoas a possibilidade de terem despesas acima da mediana aumenta 5 vezes face aos agregados que são compostos por 1 ou duas pessoas. Este resultado é semelhante ao obtido para Portugal.
- Quando os agregados familiares residem em áreas mediamente urbanas, a possibilidade de terem despesas acima da mediana é menor 10% face aos agregados familiares residentes em áreas predominantemente urbanas.
- Quando os agregados familiares residem em áreas predominantemente rurais, a possibilidade de terem despesas acima da mediana é menor 40% face aos agregados familiares residentes em áreas predominantemente urbanas.

- Quando os residentes são proprietários do alojamento e possuem crédito à habitação, a possibilidade de terem despesas acima da mediana é duas vezes superior face aos residentes que são proprietários sem crédito à habitação.
- Quando os alojamentos são constituídos por arrendatários, a possibilidade de terem despesas acima da mediana diminui para quase metade face aos residentes que são proprietários sem crédito à habitação.
- Quando os residentes ocupam alojamentos cedidos gratuitamente ou a título de salário, a possibilidade de terem despesas acima da mediana diminui 20% face aos residentes que são proprietários sem crédito à habitação.

Começámos por ajustar um modelo com todas as variáveis e posteriormente averiguamos a existência de interacções a serem incluídas. Do processo interactivo da procura do melhor modelo resultou neste caso o mesmo modelo que o indicado através do método de selecção via AIC. Na Tabela 26 podemos observar os valores ajustados e respectivas significâncias.

Tabela 26: MRL final, para Espanha

Efeito	β	Desvio padrão	Wald	valor de p	Exp(β)	95% IC Exp(β)	
						Limite inferior	Limite superior
RendTotal(1)	2,014	,058	1222,542	,000	7,495	6,695	8,391
GUrb			127,127	,000			
GUrb(1)	-,249	,062	16,271	,000	,780	,691	,880
GUrb(2)	-,605	,054	126,874	,000	,546	,491	,607
RegOcup			92,452	,000			
RegOcup(1)	,345	,076	20,432	,000	1,412	1,216	1,640
RegOcup(2)	-,769	,109	50,129	,000	,463	,374	,573
RegOcup(3)	-,173	,157	1,219	,269	,841	,618	1,144
NPAgr(1)	1,551	,058	724,247	,000	4,715	4,212	5,279
NPAgr * RegOcup			30,604	,000			
NPAgr(1) by RegOcup(1)	-,443	,084	27,808	,000	,642	,544	,757
NPAgr(1) by RegOcup(2)	-,355	,124	8,122	,004	,702	,550	,895
NPAgr(1) by RegOcup(3)	-,172	,163	1,109	,292	,842	,612	1,159
GUrb * RegOcup			34,205	,000			
GUrb(1) by RegOcup(1)	,368	,101	13,278	,000	1,445	1,185	1,761
GUrb(1) by RegOcup(2)	,214	,150	2,050	,152	1,239	,924	1,661
GUrb(1) by RegOcup(3)	-,015	,208	,005	,943	,985	,655	1,481
GUrb(2) by RegOcup(1)	,490	,093	27,471	,000	1,632	1,359	1,960
GUrb(2) by RegOcup(2)	,308	,149	4,270	,039	1,360	1,016	1,821
GUrb(2) by RegOcup(3)	,312	,179	3,031	,082	1,366	,962	1,941
NPAgr(1) by RendTotal(1)	-,396	,073	29,429	,000	,673	,583	,777
Constant	-1,507	,048	983,698	,000	,222		

3.2.1 Bondade do Ajustamento e Diagnóstico

Para a análise dos dados de Espanha, não existiam covariáveis contínuas. Com o teste de Hosmer e Lemeshow obtendo-se um valor da estatística qui-quadrado igual a 6,73, o que para 8 graus de liberdade permite aceitar a hipótese de bom ajustamento, uma vez que tem associado um valor p igual a 0,57.

Através da representação gráfica dos resíduos podemos procurar a existência de observações individuais que possam, de alguma forma, ter influência sobre a estimação dos parâmetros ou de observações inconsistentes (*outliers*).

Do gráfico da Figura 21 que representa os resíduos *deviance* nada há de saliente a registar.

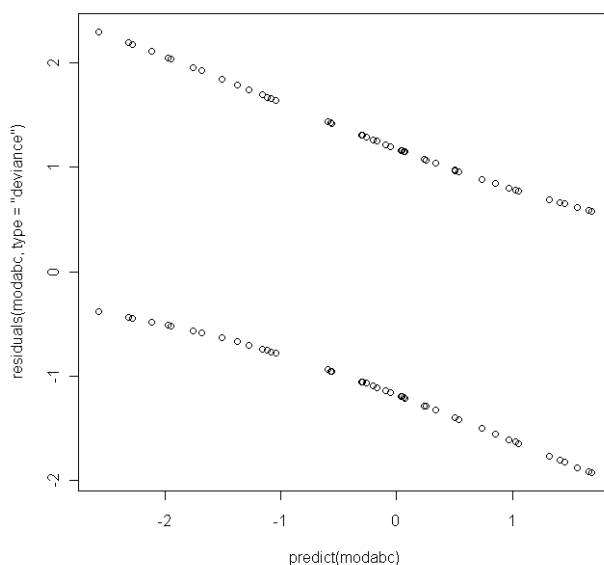


Figura 21: Resíduos *Deviance* para o modelo de Espanha

Na Figura 22 representaram-se os resíduos DFBetas. Apenas para o primeiro coeficiente se destaca uma observação que é a 112.

Representámos também a distância de Cook (Figura 23) não havendo nenhum ponto que se destaque.

Também não se registou nenhuma observação com uma repercussão muito superior às outras.

À semelhança do que fizemos no caso de Portugal, calculámos a sensibilidade e especificidade do modelo para alguns pontos de corte. Os valores estão representados na Tabela 27.

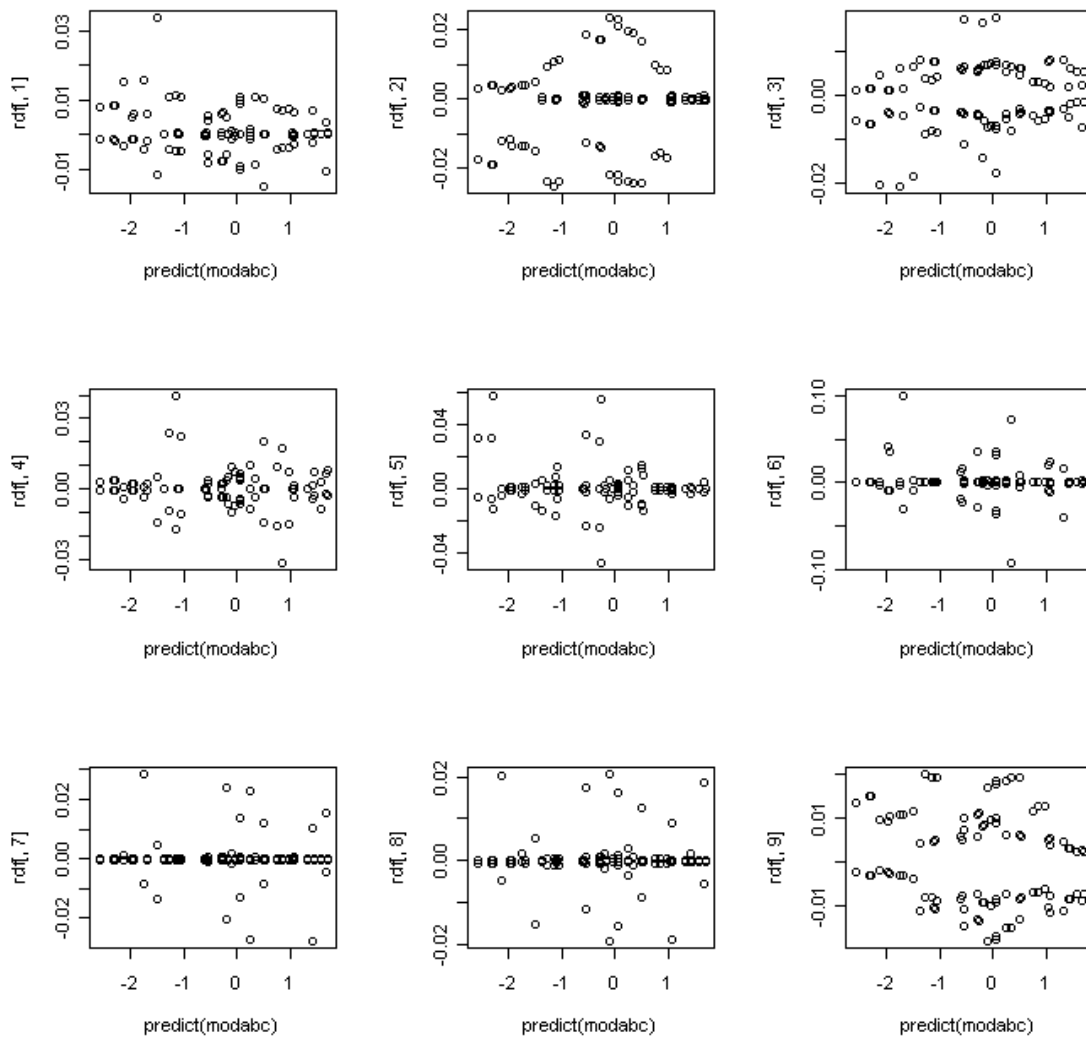


Figura 22: Resíduos Dfbetas para o modelo de Espanha

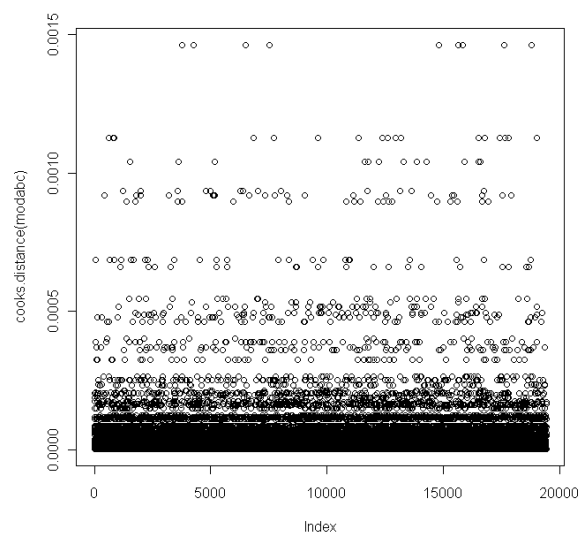


Figura 23: Distância de Cook para o Modelo de Espanha

Tabela 27: Valores de sensibilidade e especificidade obtidos, com diferentes pontos de corte para o modelo de Portugal

Quando o ponto de corte é	0,48	0,5	0,52
sensibilidade	79%	77%	72%
especificidade	71%	73%	78%

Na Figura 24 representámos a Curva ROC obtida. Este modelo tem uma capacidade classificatória também boa, mas inferior à obtida com o modelo anterior para Portugal. Neste modelo, a área estimada abaixo da curva foi igual a 0.818, estando, para um nível de confiança de 95%, entre 0.812 e 0.824.

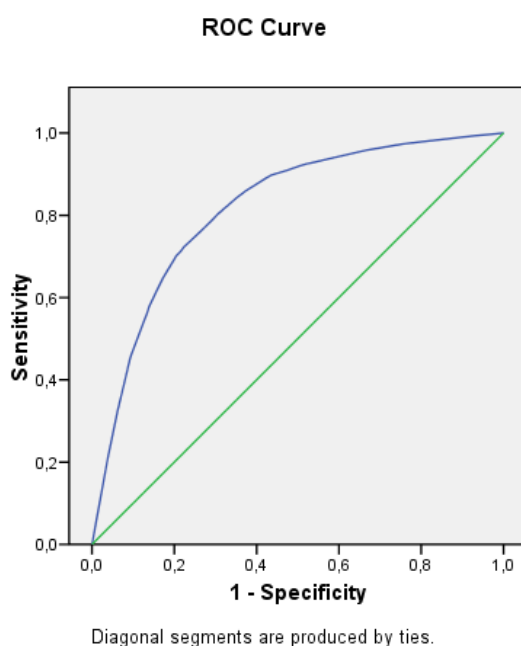


Figura 24: Curva ROC para o modelo logístico das despesas dos agregados espanhóis

3.2.2 Interpretação dos Coeficientes

A partir dos coeficientes do modelo obtido (Tabela 7) é possível retirar um conjunto de conclusões usando um raciocínio análogo ao explicado para o modelo de Portugal. Desta vez existem mais interações significativas e não foi possível simplificar mais o modelo sem que perder poder explicativo. Destacamos as seguintes conclusões:

- Para os agregados com rendimento total acima da mediana, caso o número de pessoas do agregado seja 3 ou mais, as possibilidades de terem uma despesa acima mediana aumenta cerca de 3 vezes (IC95%: 2,6; 3,8) relativamente a um agregado com 1 ou 2 pessoas (6 no caso de Portugal) enquanto para os agregados com rendimento abaixo da mediana a mesma razão de possibilidades aumenta quase 5 vezes (IC95%: 4,2; 5,3).
- Para um agregado com 3 ou mais pessoas, a possibilidade de agregados proprietários com crédito à habitação terem uma despesa acima da mediana diminui 10% (IC95%: 0,898; 0,914) relativamente a agregados sem crédito à habitação, enquanto nos agregados com 1 ou duas pessoas essa possibilidade aumenta cerca de 40% (IC95%: 1,22; 1,64).
- Para um agregado com 3 ou mais pessoas, a possibilidade de agregados arrendatários terem uma despesa acima da mediana diminui 2/3 (IC95%: 0,319; 0,331) relativamente a agregados sem crédito à habitação, enquanto nos agregados com 1 ou duas pessoas essa possibilidade diminui para aproximadamente metade (IC95%: 0,37; 0,57).
- Para um agregado proprietário com crédito à habitação a possibilidade de ter uma despesa acima da mediana quando reside numa área medianamente urbana aumenta 13% (IC95%: 1,11; 1,14) relativamente a um agregado que reside numa área predominantemente urbana, enquanto nos agregados sem crédito à habitação essa possibilidade diminui 22% (IC95%: 0,69; 0,88).
- Para um agregado proprietário com crédito à habitação a possibilidade de ter uma despesa acima da mediana quando reside numa área predominantemente rural diminui 11% (IC95%: 0,87; 0,91) relativamente a um agregado que reside numa área predominantemente urbana, enquanto nos agregados sem crédito à habitação essa possibilidade diminui para quase metade (IC95%: 0,49; 0,61).
- Para um agregado arrendatário a possibilidade de ter uma despesa acima da mediana quando reside numa área predominantemente rural diminui 25% (IC95%: 0,72; 0,77) relativamente a um agregado que reside numa área predominantemente urbana.

3.3 Modelo de Regressão Logística para Portugal e Espanha

Consideramos agora os valores da despesa categorizada em que o evento de interesse é determinado agregado familiar português ou espanhol ter um rendimento acima da mediana. Na Tabela 28 podemos observar o ajustamento do modelo com

cada variável individualmente, onde podemos constatar a significância de todas as variáveis.

Tabela 28: Resultados do ajustamento de modelos com cada variável individualmente (Portugal+Espanha)

Efeito	β	Desvio padrão	Wald	valor de p	Exp(β)	95% IC Exp(β)	
						Limite inferior	Limite superior
RendTotal(1)	2,37	0,03	7526	<0,001	10,73	10,17	11,32
País(1)	0,00	0,02	0	0,996	1,00	0,95	1,05
NPAgr(1)	1,64	0,03	4198	<0,001	5,14	4,89	5,40
RegOcup			1646	<0,001			
RegOcup(1)	0,94	0,03	1051	<0,001	2,57	2,43	2,72
RegOcup(2)	-0,60	0,04	246	<0,001	0,55	0,51	0,59
RegOcup(3)	-0,24	0,05	21	<0,001	0,78	0,71	0,87
GUrb			488	<0,001			
GUrb(1)	-0,19	0,03	40	<0,001	0,83	0,78	0,88
GUrb(2)	-0,62	0,03	488	<0,001	0,54	0,51	0,57

Tabela 29: MRL final, para Portugal e Espanha

Efeito	β	Desvio padrão	Wald	valor de p	Exp(β)	95% I.C. EXP(β)	
						Limite inferior	Limite superior
País(1)	.06	.066	.90	< 0.342	1.06	.94	1.21
RendTotal(1)	2.55	.061	1727.55	<0.000	12.75	11.31	14.38
GUrb			118.59	<0.000			
GUrb(1)	-.41	.073	31.14	<0.000	.66	.58	.77
GUrb(2)	-.79	.076	110.22	<0.000	.45	.39	.52
RegOcup			80.10	<0.000			
RegOcup(1)	.58	.085	46.81	<0.000	1.78	1.51	2.11
RegOcup(2)	-.33	.088	14.09	<0.000	.72	.60	.85
RegOcup(3)	-.22	.138	2.44	<0.118	.81	.62	1.06
NPAgr(1)	1.44	.065	488.41	<0.000	4.21	3.71	4.78
País(1) by	-.44	.064	48.68	<0.000	.64	.57	.73
NPAgr(1) by País(1)	-.18	.065	7.86	<0.005	1.20	1.06	1.36
GUrb * País			8.03	<0.018			
GUrb(1) by País(1)	-.18	.082	4.88	<0.027	1.20	1.02	1.41
GUrb(2) by País(1)	-.19	.083	5.37	<0.021	1.21	1.03	1.43
País * RegOcup			29.36	<0.000			
País(1) by	-.20	.081	6.06	<0.014	.82	.70	.96
País(1) by	-.51	.098	27.21	<0.000	.60	.50	.73
País(1) by	-.02	.135	.01	<0.911	.99	.76	1.28
NPAgr(1) by	-.54	.060	81.49	<0.000	.58	.52	.66
NPAgr * RegOcup			41.22	<0.000			
NPAgr(1) by	-.44	.073	37.19	<0.000	.64	.56	.74
NPAgr(1) by	-.30	.093	10.17	<0.001	.74	.62	.89
NPAgr(1) by	-.23	.130	3.26	<0.070	.79	.61	1.02
GUrb * RegOcup			38.02	<0.000			
GUrb(1) by	.26	.087	9.08	<0.003	1.30	1.10	1.54
GUrb(1) by	.25	.124	4.23	<0.040	1.29	1.01	1.64
GUrb(1) by	.12	.163	.57	<0.448	1.13	.82	1.56
GUrb(2) by	.42	.084	25.45	<0.000	1.53	1.30	1.80
GUrb(2) by	.40	.130	9.64	<0.002	1.50	1.16	1.93
GUrb(2) by	.47	.150	9.65	<0.002	1.59	1.19	2.14

A partir dos resultados contidos podemos tecer as seguintes considerações:

- Quando os agregados familiares auferem um total de rendimentos acima da mediana, a possibilidade de terem despesas acima da mediana aumenta cerca de 10 vezes face aos agregados que auferem rendimentos abaixo da mediana.
- Quando o número de indivíduos do agregado é composto por 3 ou mais pessoas a possibilidade de terem despesas acima da mediana aumenta 5 vezes face aos agregados que são compostos por 1 ou duas pessoas.

- Para os residentes que são proprietários do alojamento e possuem crédito à habitação a possibilidade de terem despesas acima da mediana aumenta 2 vezes e meia face aos residentes que são proprietários sem crédito à habitação.
- Quando os alojamentos são constituídos por arrendatários, a possibilidade de terem despesas acima da mediana diminui para metade face aos residentes que são proprietários sem crédito à habitação.
- Quando os residentes ocupam alojamentos cedidos gratuitamente ou a título de salário, a possibilidade de terem despesas acima da mediana diminui 13% face aos residentes que são proprietários sem crédito à habitação.
- Quando os agregados familiares residem em áreas mediantemente urbanas, a possibilidade de terem despesas acima da mediana diminui 12% face aos agregados familiares residentes em áreas predominantemente urbanas.
- Quando os agregados familiares residem em áreas predominantemente rurais, a possibilidade de terem despesas acima da mediana é metade da dos agregados familiares residentes em áreas predominantemente urbanas.
- Quando os agregados familiares pertencem a Espanha, a possibilidade de terem despesas acima da mediana não difere significativamente dos agregados familiares que pertencem a Portugal.

Começámos por ajustar um modelo com todas as variáveis, incluindo o país, apesar de não ser individualmente significativo e posteriormente averiguámos a existência de interacções a serem incluídas. Do processo interactivo na procura do melhor modelo resultou neste caso um modelo mais parcimonioso que um outro que apresentava mais interacções, mas que quando testado usando a razão de verosimilhanças se verificou não ser significativamente melhor. Na Tabela 28 podemos observar os valores ajustados e respectivas significâncias.

3.3.1 Bondade do Ajustamento e Diagnóstico

Como estes são apenas dos dois países e dado que, para Espanha, não foi possível apurar o índice de conforto, também neste caso temos um modelo sem covariáveis contínuas. Com o teste de Hosmer e Lemeshow obtendo-se um valor da estatística qui-quadrado igual a 8.53, o que, para 8 graus de liberdade, permite aceitar a hipótese de bom ajustamento, uma vez que tem associado um valor p igual a 0,38.

Através da representação gráfica dos resíduos podemos procurar a existência de observações individuais que possam, de alguma forma, ter influência sobre a estimação dos parâmetros ou de observações inconsistentes (*outliers*).

Do gráfico da Figura 25 que representa os resíduos *deviance* nada há de saliente a registar.

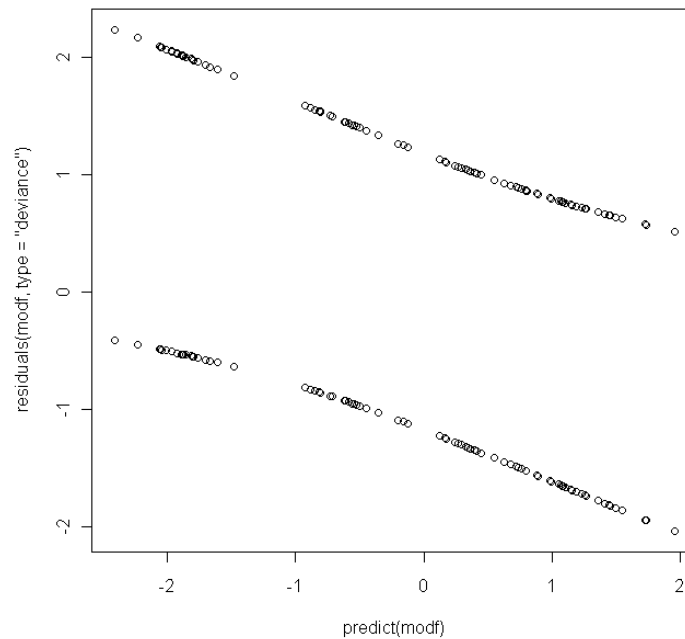


Figura 25: Resíduos *Deviance* para o modelo conjunto

Na Figura 26 representaram-se os resíduos DFBetas. Apenas para o primeiro coeficiente se destaca uma observação que é a número 19498. De acordo com a disposição da base de dados, é uma observação relativa a Portugal.

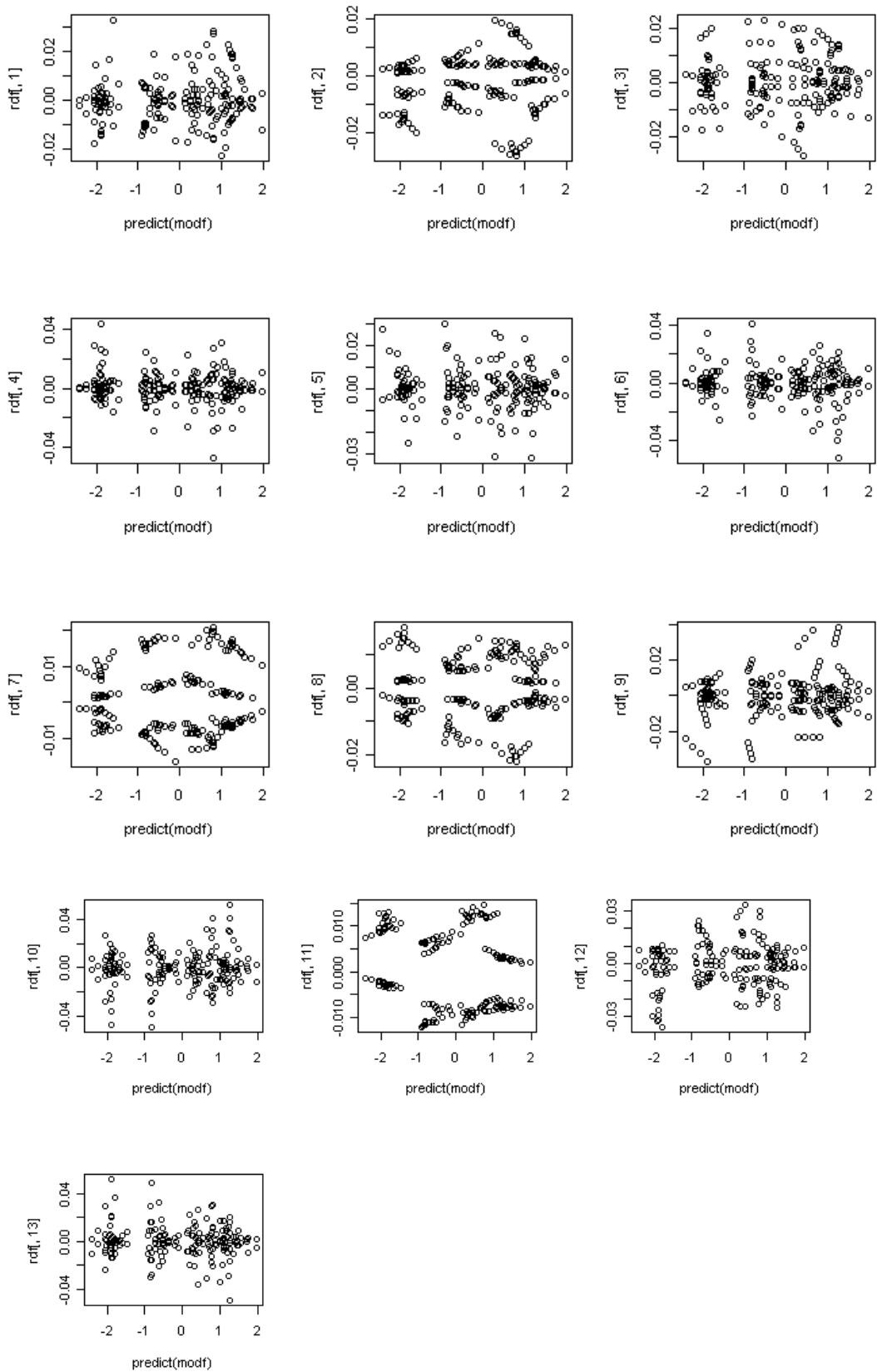


Figura 26: Resíduos Dfbetas para o modelo conjunto

Representámos também a distância de Cook (Figura 27) e embora se destaquem as observações 24591, 25525, 27374 e 28253, todas são relativas a Portugal, o valor é demasiado pequeno para terem uma grande importância na estimação dos parâmetros, o que foi confirmado ao ajustar o modelo sem estas observações.

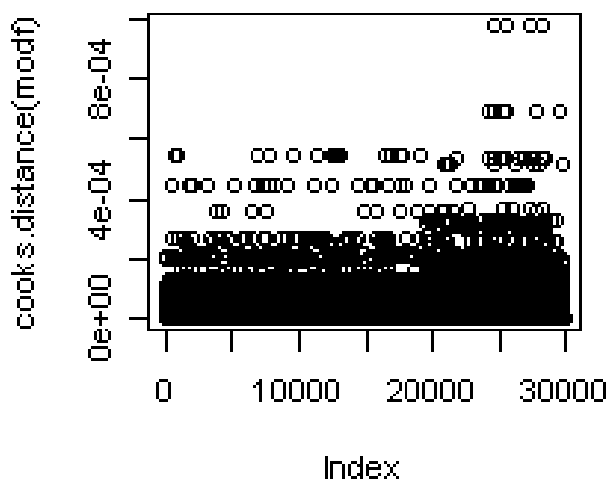


Figura 27: Distância de Cook para o Modelo de Portugal e Espanha

Neste caso o ponto de corte que maximiza a sensibilidade e especificidade é 0,53, resultando num modelo com 79% de sensibilidade e 74% de especificidade.

Na Figura 28 representámos a Curva ROC obtida. Este modelo tem uma capacidade classificatória também boa, mas inferior à obtida com o modelo para Portugal e ligeiramente superior à obtida com o modelo para Espanha. Neste modelo, a área estimada abaixo da curva foi igual a 0.832, estando, para um nível de confiança de 95%, entre 0.828 e 0.837.

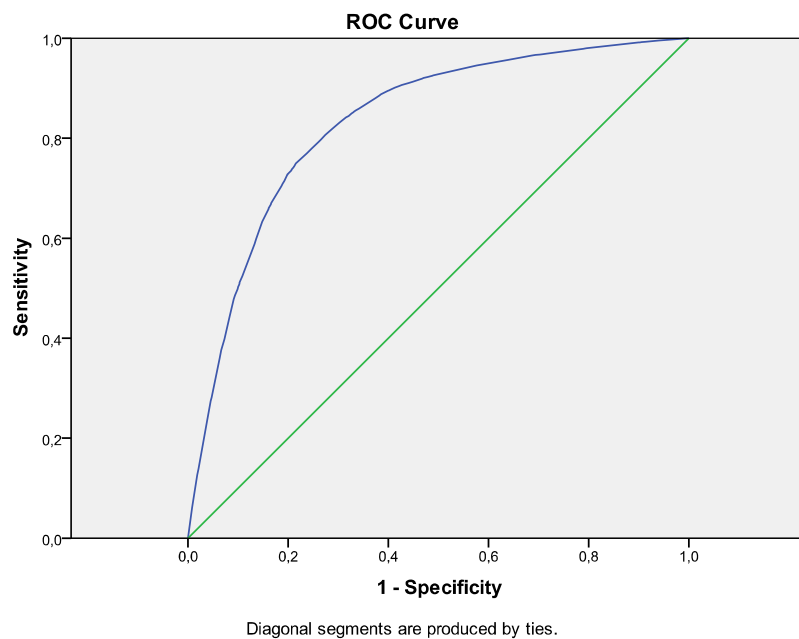


Figura 28: Curva ROC para o modelo logístico das despesas dos agregados portugueses e espanhóis.

3.3.2 Interpretação dos Coeficientes

A partir dos coeficientes do modelo obtido (Tabela 28) é possível retirar um conjunto de conclusões usando um raciocínio análogo ao explicado para o modelo de Portugal. Desta vez também existem muitas interações significativas. Deixamos a interpretação da variável país neste modelo que permite comparar os dois países. Assim, os agregados espanhóis têm 10% menos de possibilidades de terem uma despesa acima da mediana que os agregados portugueses.

Capítulo IV

4. Considerações finais

Apesar de ambos os países estarem integrados na União Europeia, não existia em 2005-2006 uma total harmonização em operações estatísticas comuns, que tinham como objectivo recolher informação de modo a que a mesma fosse conduzida a nível comunitário. Quando se pretende caracterizar e comparar a informação obtida entre países que partilham o mesmo espaço comunitário, são identificadas muitas dificuldades que têm início logo no acesso à informação dos microdados, quando no caso de Espanha se encontram à distância de um *click* (em <http://www.ine.es/prodyser/microdatos.htm>), em Portugal é necessário pedir, justificar, assinar, carimbar e...aguardar. O facto do desenho do questionário apresentar diferenças entre os países vem realçar as prováveis desigualdades que culminam nas poucas variáveis que ambos os países apresentam em comum. No entanto, as dificuldades encontradas converteram-se em desafios.

Através da análise exploratória foi possível caracterizar de forma simples os agregados familiares portugueses, ficando a saber que, em 2005/2006, acima de 50% dos casos, os agregados familiares portugueses residiam em alojamentos dos quais eram proprietários sem que para isso recorressem ao crédito à habitação, cerca de 50% dos agregados familiares eram constituídos por uma ou duas pessoas e mais de 60% dos agregados familiares inquiridos residiam em áreas predominantemente urbanas. E identificar diferenças que intuitivamente já existiam, (ex. a média do rendimento total é mais elevada em Espanha, a amplitude dos valores do rendimento total é mais elevada em Portugal, é a despesa monetária que representa a maior proporção na despesa total em ambos os países, as médias das despesas total e monetária são mais elevadas em Espanha) e foi possível a obtenção de indicadores menos prováveis (ex. é em Portugal que se verifica o valor mais elevado de rendimento total) acima de 50% dos casos, os agregados familiares portugueses residiam em alojamentos dos quais eram proprietários sem que para isso recorressem ao crédito à habitação).

Através do estudo de medidas de concentração foi possível analisar os graus de concentração dos rendimentos e das despesas, sendo que se obteve uma muito maior concentração na distribuição dos rendimentos não monetários do que dos rendimentos monetários em Portugal. Concluiu-se também num maior índice de concentração das

despesas em Espanha do que em Portugal, sendo a concentração dos rendimentos muito idênticas. No entanto, verificou-se que na distribuição dos rendimentos e das despesas mais elevadas (em especial, os 5% e 10% mais elevados) se regista uma maior diferença entre os dois países, sendo os valores mais elevados obtidos para Portugal.

O processo de modelação estatística proporcionou-nos a utilização de dois softwares estatísticos (SPSS e R), em momento diferentes.

Através da categorização foi possível a verificação das possíveis relações existentes entre as variáveis com o objectivo de conseguir ir ainda mais além das análises anteriormente feitas e publicadas e, simultaneamente, de poder comparar os dois países em termos da explicação das despesas dos agregados familiares por variáveis comuns. Foram sendo construídos, ajustados, verificados pressupostos e comparados modelos, através da regressão logística. Os modelos obtidos apresentam uma muito boa capacidade classificatória.

As principais conclusões retiradas na fase de modelação foram sendo deixadas ao longo do texto. Deixamos, no entanto, uma ou outra que nos parece de destacar.

Quando os agregados familiares Portugueses auferem um total de rendimentos não monetário acima da mediana, a possibilidade de terem despesas acima da mediana aumenta 4 vezes face aos agregados que auferem rendimento não monetário abaixo da mediana.

Ainda para os agregados Portugueses, quando os alojamentos são constituídos por arrendatários, a possibilidade de terem despesas acima da mediana aumenta 23% face aos residentes que são proprietários sem crédito à habitação, diminuindo 6% se os residentes ocupam alojamentos cedidos gratuitamente ou a título de salário.

Para os agregados com rendimento monetário acima da mediana, caso o número de pessoas do agregado seja três ou mais, as possibilidades de terem uma despesa acima da mediana aumenta 6 vezes se é Português e 3 vezes se é Espanhol relativamente a agregados com uma ou duas pessoas, enquanto para os agregados com rendimento abaixo da mediana a mesma razão de possibilidades aumenta 9 vezes se é Português e quase 5 se é Espanhol.

Para um perfil fixo relativamente às variáveis regime de ocupação, número de pessoas do agregado, rendimento total e grau de urbanização, os agregados espanhóis têm 10% menos de possibilidades de terem uma despesa acima da mediana que os agregados portugueses.

A análise apresentada neste trabalho mostra que é possível fazer “falar” mais os dados do que eles anteriormente tinham “falado”. Pensamos que outras abordagens são possíveis e desejáveis e a comparação destes resultados com os do inquérito do próximo ano será um passo inevitável.

5. Bibliografia

Referências bibliográficas

- [1] Agresti, A. (2007). *An Introduction to categorical data analysis*, Second Edition, Wiley (p. 66-80).
- [2] Harrell, L. and Mark K. L. (1996). *Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors*, Stat in Med, (p. 361-387).
- [3] Hosmer, D. W. and Lemeshow, S. (2000). *Applied Logistic Regression*, John Wiley, New York, (p. 3-175).
- [4] Lorenz, M. O. (1905). *Methods of measuring the concentration of wealth*, Pub. 9 of the American Statistical Association (p. 209-219).
- [5] McCulloch, C. E. and Nelder, J.A. 1989). *Generalized linear models*, second edition; Chapman and Hall, London.
- [6] McCulloch, C. E. and Searle, S. R. (2001) *Generalized, linear and mixed models*. J. Wiley and Sons (p. 135-160).
- [7] Murteira, B. e Black, H. (1983). *Estatística Descritiva*, McGraw-Hill, (p. 100-120).
- [8] Paula, G. A. (2004), Modelos de regressão com apoio computacional. Instituto de Matemática e estatística da Universidade de S. Paulo, (p. 28-96).
- [9] Turkman, M. A. Amaral e Silva, G. (2000). *Modelos Lineares Generalizados – da Teoria à Prática*, Edições SPE, Lisboa, (p. 3-98).

Outras referências:

1. www.ats.ucla.edu
2. www.cso.ie Central Statistics Office Ireland;
3. <http://epp.eurostat.ec.europa.eu>, European Commission Eurostat;
4. Hauck, W. J. and Donner, A. Journal of the American Statistical Association. Vol. 72, N.º 360 (1977), pp 851-853. www.jstor.org
5. www.imf.org International Monetary Fund;
6. www.ine.es, Instituto Nacional de Estadística (Espanha);
7. Inquérito às Despesas das Famílias 2005-2006, INE (2008), Portugal, www.ine.pt;
8. Nelder, J. A. and Wedderburn, R. W. M., Generalized Linear Models, *Journal of the Royal Statistical Society. Series A (General)*. Vol. 135, N.º 3 (1972), pp. 370-384 [Blackwell Publishing](http://www.blackwellpublishing.com) for the [Royal Statistical Society](http://www.royalsocietypublishing.org) www.jstor.org.

Anexos

Anexo I

Questionário utilizado no Inquérito às Despesas das Famílias (2005/2006)

Instrumento de Notação do Sistema Estatístico Nacional
(Lei n.º 836 de 15 de Abril), de natureza obrigatória
Registo no I.N.E. sob o nº 9886
Válido até 31/10/2008



INSTITUTO NACIONAL DE ESTATÍSTICA
PORTUGAL

INQUÉRITO ÀS DESPESAS DAS FAMÍLIAS

- 2005 / 2006 -

MÓDULO I

CARACTERIZAÇÃO DO ALOJAMENTO, DO AGREGADO E DO(S) INDIVÍDUO(S)

ENQUADRAMENTO DO INQUÉRITO ÀS DESPESAS DAS FAMÍLIAS (IDEF)

O Instituto Nacional de Estatística (INE) realiza inquéritos aos orçamentos familiares desde 1967, com o objectivo de avaliar as condições de vida dos indivíduos residentes em Portugal, bem como conhecer como são aplicados os seus rendimentos em despesas de consumo.

O actual quadro legal (comunitário) estabelece a necessidade de se realizarem estudos periódicos sobre esta realidade, pelo que o INE vem solicitar e agradecer a vossa colaboração enquanto família seleccionada para responder ao inquérito no contexto de um processo de amostragem aleatória.

Salienta-se o seguinte :

Todos os dados individuais recolhidos no âmbito deste inquérito são confidenciais, estando protegidos por segredo estatístico (art.º 5 da Lei nº6/89 de 15 de Abril), e destinando-se apenas à obtenção de resultados globais sobre o consumo dos agregados.

O INE assegura a destruição dos questionários preenchidos, após a obtenção e divulgação de resultados.

O Entrevistador do INE, devidamente credenciado, que periodicamente o visitará, prestará todos os esclarecimentos que necessitar e ajudará a resolver as dúvidas que lhe surjam.

Entrevista	Data de início da quinzena <input type="text"/> / <input type="text"/> / <input type="text"/> 2 0 0 <input type="text"/>		
Entrevistador	Nome <input type="text"/>	Código	<input type="text"/>
Identificação	Área <input type="text"/>	Alojamento <input type="text"/>	Agregado <input type="text"/>
	Nome <input type="text"/>		
	Morada (rua, av, pc, apartado, etc) <input type="text"/>		
	Nº ou lote <input type="text"/> Andar, sala, etc <input type="text"/> Lado <input type="text"/>		
	Localidade <input type="text"/>		
	Código postal <input type="text"/>		
Telefone(s) para contacto <input type="text"/> ou <input type="text"/>			

MÓDULO I

I.1 - ALOJAMENTO

I.1.1	Situação do alojamento / Resultado do contacto	<input type="text"/>
	Residência principal - Entrevista conseguida	1
	Residência principal - Temporariamente ausente	2
	Residência principal - Recusa	3
	Residência secundária	4
	Alojamento vago	5
	Alojamento inlocalizável	6
	Alojamento demolido	7
	Outre situação:	8
	<div style="border: 1px solid black; padding: 2px; display: inline-block;">Outre situação:</div> Especifique, por favor: _____ _____ _____	
I.1.2	Tipo de alojamento	<input type="text"/>
	Moreia independente isolada	1
	Moreia independente geminada ou em banda	2
	Apartamento num edifício com menos de 10 apartamentos	3
	Apartamento num edifício com 10 ou mais apartamentos	4
	Barraco	5
	Outro tipo de alojamento	6
I.1.3	1. Ano de construção do alojamento	<input type="text"/>
		se não sabe => I.1.3.2
	2. Década de construção - 1º ano (antes de 1900: 1890; ...; anos 20: 1920; ...; anos 70: 1970; ...)	<input type="text"/>
I.1.4	Número de agregados a residir no alojamento	<input type="text"/>

I.2 - AGREGADO

I.2.1	Número de identificação do agregado	<input type="text"/>
I.2.2	Regime de ocupação do alojamento	<input type="text"/>
	Proprietário	sem crédito à habitação 1 com crédito à habitação 2
	Arrendatário (ou subarrendatário)	com renda a preços de mercado 3 com renda inferior ao preço de mercado 4
	Alojamento cedido gratuitamente ou a título de salário	5
I.2.3	Nº divisões disponíveis para o agregado (4 m ² ou +)	<input type="text"/>
I.2.4	Área total (m²) disponível para o agregado (espaço útil entre paredes)	<input type="text"/>
I.2.5	Disponibilidade de bens	
	I.2.5.1 Geregem (ou espaço para parqueamento) na residência principal?	<input type="text"/> 1: Sim
	I.2.5.2 Residência secundária?	<input type="text"/> 2: Não
	I.2.5.2.1 Quantas resid. secund.?	<input type="text"/>
I.2.6.2	Regime de ocupação da(s) residência(s) secundária(s)	
	I.2.6.2.a Nº de residências secundárias que o agregado dispõe - Proprietário	<input type="text"/>
	I.2.6.2.b Nº de residências secundárias que o agregado dispõe - Arrendatário	<input type="text"/>
	I.2.6.2.c Nº de residências secundárias que o agregado dispõe - Cedida gratuitamente ou a título de salário	<input type="text"/>

I.2.6 - CONFORTO E BENS DE EQUIPAMENTO

Conforto básico no interior do alojamento	Equipamento de comunicação, áudio e vídeo			
I.2.6.1 Água canalizada <input type="checkbox"/>	I.2.6.21 Telefone - rede fixe <input type="checkbox"/>			
I.2.6.2 Electricidade <input type="checkbox"/>	I.2.6.22 Telefone - rede móvel <input type="checkbox"/> <small>se Não, I.2.6.23</small> Quantos? <input type="checkbox"/> (1-10)			
I.2.6.3 Gás canalizado <input type="checkbox"/>	I.2.6.23 Aparelho de televisão <input type="checkbox"/> <small>se Não, I.2.6.24</small> Quantos? <input type="checkbox"/> (1-10)			
I.2.6.4 Sistema de esgotos <input type="checkbox"/>	I.2.6.24 Antena parabólica <input type="checkbox"/>			
I.2.6.5 Instalação sanitária completa <input type="checkbox"/>	I.2.6.25 Televisão por cabo ou satélite <input type="checkbox"/>			
Equipamento de apoio ao trabalho doméstico				
I.2.6.6 Fogão ou placa <input type="checkbox"/>	I.2.6.26 Leitor de CD's <input type="checkbox"/>			
I.2.6.7 Micro-ondas <input type="checkbox"/>	I.2.6.27 Leitor de DVD's <input type="checkbox"/>			
I.2.6.8 Frigorífico <input type="checkbox"/>	I.2.6.28 Gravador de cassetes áudio <input type="checkbox"/>			
I.2.6.9 Arca congeladora <input type="checkbox"/>	I.2.6.29 Rádio <input type="checkbox"/>			
I.2.6.10 Aspirador <input type="checkbox"/>	I.2.6.30 Gire-Discos <input type="checkbox"/>			
I.2.6.11 Máquina de lavar roupa <input type="checkbox"/>	I.2.6.31 Vídeo <input type="checkbox"/>			
I.2.6.12 Máquina de secar roupa <input type="checkbox"/>	I.2.6.32 Câmara de vídeo <input type="checkbox"/>			
I.2.6.13 Máquina de lavar e secar roupa <input type="checkbox"/>	I.2.6.33 Equipamento fotográfico <input type="checkbox"/>			
I.2.6.14 Máquina de lavar loiça <input type="checkbox"/>	I.2.6.34 Computador pessoal, de secretária ou portátil, COM ligação à internet <input type="checkbox"/>			
I.2.6.15 Máquina de costura <input type="checkbox"/>	I.2.6.35 Computador pessoal, de secretária ou portátil, SEM ligação à internet <input type="checkbox"/>			
<div style="display: flex; justify-content: space-around; margin-top: 10px;"> 1: Sim 2: Não </div>				
Equipamento para regular o ar e a água				
I.2.6.16 Aparelho de ar condicionado <input type="checkbox"/>	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 10%; text-align: center;">I.2.7</td> <td style="width: 60%;">Número de indivíduos membros do agregado</td> <td style="width: 30%; text-align: center;"><input type="checkbox"/></td> </tr> </table>	I.2.7	Número de indivíduos membros do agregado	<input type="checkbox"/>
I.2.7		Número de indivíduos membros do agregado	<input type="checkbox"/>	
I.2.6.17 Sistema de aquecimento central <input type="checkbox"/>				
I.2.6.18 Outro aparelho de aquecimento de ar <input type="checkbox"/>				
I.2.6.19 Outro aparelho de aquecimento de água <input type="checkbox"/>				
I.2.6.20 Desumidificador eléctrico <input type="checkbox"/>				

I.3 - CARACTERIZAÇÃO DO(S) INDIVÍDUO(S)

I.3.1 Identificação do indivíduo					
.01	Nº ordem	_	_	_	_
.02	Nome				

I.3.2	Data de nascimento	(dd/mm/aaaa)	_ / _ /	_ / _ /	_ / _ /	_ / _ /
--------------	---------------------------	--------------	---------	---------	---------	---------

I.3.3	Sexo	masculino: 1	feminino: 2	_	_	_	_
--------------	-------------	--------------	-------------	---	---	---	---

I.3.4	Situação residencial	presente: 1	temporária/ausente: 2	_	_	_	_
--------------	-----------------------------	-------------	-----------------------	---	---	---	---

I.3.5	Relação com o representante		_	_	_	_
	O próprio representante: 1	Filho(a) do representante ou do cônjuge: 3	Outro parente: 5			
	Cônjuge/companheiro(a): 2	Pai/Mãe do representante ou do cônjuge: 4	Sem parentesco: 6			

I.3.6	Nível de escolaridade completado		_	_	_	_
	Nenhum: 1	Básico – 3º ciclo (9º ano): 4	Superior – licenciatura: 7			
	Básico – 1º ciclo (4ª ano/idade): 2	Secundário (12º ano) e pós-secundário: 5	Superior – mestrado: 8			
	Básico – 2º ciclo (6º ano): 3	Superior – bacharelato: 6	Superior – doutoramento: 9			

I.3.7	Condição perante o trabalho habitual		_	_	_	_
	Exerce uma profissão: 1	Aluno ou estudante: 4	Incapacitado permanentem. para o trabalho: 7			
	Desempregado: 2	Serviço cívico: 5	Outros inactivos (incluindo pensionistas): 8			
	Reformado/aposentado: 3	Doméstico: 6				

Se tem menos de 15 anos e I.3.7 -> 1 (não exerce profissão) -> FIM

I.3.8	Situação na profissão		_	_	_	_
	Trabalhador por conta própria - empregador: 1	Trabalhador por conta de outrem: 3	Aprendizes ou estagiários: 5			
	Trabalhador por conta própria - isolado: 2	Trabalhador familiar não remunerado: 4	Outros: 6			

I.3.9	Indique se teve receitas monetárias no ano anterior		_	_	_	_
	1: Sim	2: Não				

RECEITAS MONETÁRIAS LÍQUIDAS DO INDIVÍDUO (NO ANO ANTERIOR)		
I.3.1 Identificação do indivíduo (conforme identificação anterior)		
.01 Nº ordem	.02 Nome	
I.3.10 Indique as Receitas Monetárias Líquidas Anuais que obteve no ano anterior para as diferentes categorias apresentadas:		
Rendimentos do trabalho		
I.3.10.01	Por conta de outrem - regulares	_____ €
I.3.10.02	Por conta de outrem - ocasionais	_____ €
I.3.10.03	Por conta própria	_____ €
Rendimentos de Propriedade		
I.3.10.04	Rendas de terrenos	_____ €
I.3.10.05	Rendas de edifícios	_____ €
Rendimentos de Capital		
I.3.10.06	Juros	_____ €
I.3.10.07	Lucros e dividendos	_____ €
I.3.10.TPP	Transferências Periódicas: Pensões - OBTVEU NO ANO ANTERIOR? <i>(pensões de velhice, pensões de sobrevivência, pensões de invalidez e outras pensões privadas de estrangeiros)</i>	<input type="checkbox"/> 1: Sim <input type="checkbox"/> 2: Não
I.3.10.08	Pensão de velhice	_____ €
I.3.10.09	Pensão social de velhice	_____ €
I.3.10.10	Pensão de viuvez e/ou orfandade	_____ €
I.3.10.11	Pensão de sobrevivência	_____ €
I.3.10.12	Pensão de invalidez	_____ €
I.3.10.13	Outras pensões do sistema de segurança social pública	_____ €
I.3.10.14	Pensões de regimes privados de segurança social	_____ €
I.3.10.15	Pensões do estrangeiro	_____ €
I.3.10.OTP	Outras Transferências Periódicas - OBTVEU NO ANO ANTERIOR? <i>(abono de família, benefícios de alojamento, transferências periódicas de outros agregados, do estrangeiro, etc.)</i>	<input type="checkbox"/> 1: Sim <input type="checkbox"/> 2: Não
I.3.10.16	Abono de família	_____ €
I.3.10.17	Benefícios relacionados com o alojamento	_____ €
I.3.10.18	Transferências periódicas de outros agregados (pensão de alimentos, apoio dos filhos, pais, etc.)	_____ €
I.3.10.19	Transferências periódicas do estrangeiro	_____ €
I.3.10.20	Outras transferências periódicas	_____ €
I.3.10.TNP	Transferências não Periódicas - OBTVEU NO ANO ANTERIOR? <i>(benefícios à família, outros benefícios por desemprego, doença, invalidez, educação, formação, apoio social, resgate (montante garantido), transferências não regulares de outros agregados, do estrangeiro, etc.)</i>	<input type="checkbox"/> 1: Sim <input type="checkbox"/> 2: Não
I.3.10.21	Benefícios relacionados com a família (subsídio de casamento, subsídio de eleição, subsídio de adoção, etc.)	_____ €
I.3.10.22	Benefícios relacionados com desemprego (indenizações por despedimento, subsídio de desemprego, subsídio social de desemprego, compensação salarial, etc.)	_____ €
I.3.10.23	Benefícios relacionados com doença ou invalidez (subsídio de doença, subsídio de tuberculose, subsídio de acompanhamento ou assistência a 3ª pessoa, etc.)	_____ €
I.3.10.24	Benefícios relacionados com educação (subsídio de educação especial, bolsas de estudo, etc.)	_____ €
I.3.10.25	Benefícios relacionados com a formação (subsídio de reconversão profissional, subsídio de formação profissional e jovens, subsídio de formação do FSE, etc.)	_____ €
I.3.10.26	Rendimento Social de Inserção (Rendimento mínimo garantido)	_____ €
I.3.10.27	Transferências não periódicas de outros agregados (montante dado ocasionalmente por filhos/pais, etc.)	_____ €
I.3.10.28	Outras transferências não periódicas do estrangeiro	_____ €
I.3.10.29	Outras transferências não periódicas	_____ €



INSTITUTO NACIONAL DE ESTATÍSTICA
PORTUGAL

INQUÉRITO ÀS DESPESAS DAS FAMÍLIAS

- 2005 / 2006 -

MÓDULO II

DIÁRIO DE CONSUMO DO AGREGADO

[AUTO-PREENCHIMENTO DURANTE 1 QUINZENA]

ENQUADRAMENTO DO INQUÉRITO ÀS DESPESAS DAS FAMÍLIAS (IDEF)

O Instituto Nacional de Estatística (INE) realiza inquéritos aos orçamentos familiares desde 1967, com o objectivo de avaliar as condições de vida dos indivíduos residentes em Portugal, bem como conhecer como são aplicados os seus rendimentos em despesas de consumo.

O actual quadro legal (comunitário) estabelece a necessidade de se realizarem estudos periódicos sobre esta realidade, pelo que o INE vem solicitar e agradecer a vossa colaboração enquanto família seleccionada para responder ao inquérito no contexto de um processo de amostragem aleatória.

Salienta-se o seguinte :

Todos os dados individuais recolhidos no âmbito deste inquérito são confidenciais, estando protegidos por segredo estatístico (art.º 5 da Lei nº6/89 de 15 de Abril), e destinando-se apenas à obtenção de resultados globais sobre o consumo dos agregados.

O INE assegura a destruição dos questionários preenchidos, após a obtenção e divulgação de resultados.

O Entrevistador do INE, devidamente credenciado, que periodicamente o visitará, prestará todos os esclarecimentos que necessitar e ajudará a resolver as dúvidas que lhe surjam.

Entrevista	Data de início da quinzena <input type="text"/> / <input type="text"/> / <input type="text"/> 2 0 0 <input type="text"/>		
Entrevistador	Nome <input type="text"/>	Código	<input type="text"/>
Identificação	Área <input type="text"/>	Alojamento <input type="text"/>	Agregado <input type="text"/>
	Nome <input type="text"/>		
	Morada (rua, av, pc, apartado, etc) <input type="text"/>		
	Nº ou lote <input type="text"/> Andar, sala, etc <input type="text"/> Lado <input type="text"/>		
	Localidade <input type="text"/>		
	Código postal <input type="text"/>		
Telefone(s) para contacto <input type="text"/> ou <input type="text"/>			

MÓDULO II
DIÁRIO DE CONSUMO DO AGREGADO - EXCEPTO SAÚDE
(COMPRAS, RECEBIMENTOS GRATUITOS, AUTO-CONSUMO E AUTO-ABASTECIMENTO)

INSTRUÇÕES DE PREENCHIMENTO GERAIS

Deverão ser preenchidas as despesas correntes do agregado, anotando-se todos os bens e serviços consumidos, qualquer que seja o seu valor. Se algum indivíduo não quiser preencher o diário individual, poderão ser aqui incluídos os seus registos (por exemplo, dos indivíduos entre 10 e 14 anos).

Utilizar uma linha para cada bem ou serviço consumido, podendo ser agrupados apenas no caso dos bens em questão serem absolutamente iguais (como por exemplo, 8 iogurtes ou 4 embalagens de lenços de papel).

As anotações neste diário devem ser efectuadas o mais breve possível após cada consumo de um bem ou serviço, de modo a que não sejam esquecidas quaisquer situações, tornando-se assim muito mais fácil obter um preenchimento organizado e exaustivo.

Se, num determinado dia, forem efectuadas compras em grande quantidade em hipermercado ou supermercado, havendo talão de caixa onde esteja bem clara a descrição dos produtos, quantidades e preços, poderá a anotação no diário ser substituída pelo próprio talão, colado ou agramado ao diário. Cada talão deverá ser cuidadosamente lido, de modo a confirmar-se que toda a informação necessária nele esteja contida e bem clara. Sempre que determinado produto não esteja perceptível (descrição, quantidade ou valor), deve-se-á escrever no ticket tudo o que possa completar a sua correcta identificação. A data terá de ser bem legível.

INSTRUÇÕES DE PREENCHIMENTO DETALHADAS

Modalidade de consumo (colunas 2, 3, 4 e 5)	Uma das quatro colunas (2, 3, 4 e 5) necessitará de ser preenchida com X, consoante se trate de:
Compra (coluna 2)	Aplica-se a aquisições de bens e serviços em contrapartida de um pagamento (no momento ou posteriormente).
Recebimento gratuito (coluna 3)	Aplica-se aos bens e serviços recebidos sem contrapartida de pagamento (prezados, por ex.), e que não sejam um complemento ao salário. É importante não esquecer eventuais refeições recebidas vindas de fora do agregado, quer sejam de família (a viver em outros agregados), de amigos, vizinhos ou instituições de solidariedade.
Auto-consumo (coluna 4)	Refere-se a consumo de bens alimentares de produção própria do agregado (galinhas de própria criação, couves do quintal, coelhos caçados ou peixes pescados, por exemplo).
Auto-abastecimento (coluna 5)	Adquire os bens e serviços provenientes de estabelecimentos explorados por algum membro do agregado e que não sejam pagos (por exemplo, vestuário retirado de loja do agregado, ...).
Compra no estrangeiro? (coluna 6)	Esta coluna só deverá ser preenchida se a despesa ocorrer fora de Portugal. Neste caso, preenche-se com X.
Descrição do produto ou serviço (coluna 7)	Deverá ser apresentada uma descrição muito clara, seja qual for a modalidade de consumo, por forma a permitir uma correcta codificação. Reafirma-se que só se podem agrupar numa mesma linha produtos absolutamente equivalentes.
Quantidade (coluna 8)	A quantidade deverá ser expressa com a indicação da unidade de medida, sempre que possível de acordo com o explicitado na COICOP, como por exemplo, 1 kg, 2 litros, 3 metros, 1 dúzia, 1 pack com 6 unidades, 2 embalagens de 125 ml cada, 1 unidade, ... No caso dos serviços, a quantidade também deverá ser preenchida, ainda que, na grande maioria das situações, seja "1". Se, por exemplo, em determinado dia, 2 crianças do agregado forem ao pediatra, o valor indicado deverá ser a soma das duas consultas médicas, com a indicação de "2" na quantidade.
Valor (coluna 9)	O valor deverá ser expresso em Euros, utilizando os dois casos decimais para os centimos, e deve referir-se à totalidade da quantidade dos bens ou serviços indicados em cada linha. Mesmo que o pagamento não seja efectuado na ocasião, deve ser sempre preenchido o valor total respectivo. Em caso de auto-consumo, o valor a preencher será aquele que se pagaria no estabelecimento mais próximo a que recorreria. Em caso de auto-abastecimento, o valor a preencher será o praticado em caso de venda ou prestação de serviço a um cliente. Em caso de recebimentos gratuitos, o valor a preencher será aquele que se pagaria no estabelecimento mais próximo a que recorreria.
Tipo de estabelecimento ou de profissional independente (coluna 10)	Se for um estabelecimento especializado em determinado ramo, deverá ser descrito desse modo, anotando-se, por exemplo, padaria, talho, peixaria, pastelaria, restaurante, café, farmácia, tabacaria, papelaria, pronto-vestido, sapataria, ..., consultório médico, banco, ginásio, cinema (ver tabela anexa). Se for um estabelecimento não especializado, deverá ser indicado de acordo com a definição que o próprio estabelecimento atribui a ele mesmo, como por exemplo, mercearia, minimercado, super-mercado, hiper-mercado, praça, quiosque de rua, banca em feira, vendedor ambulante (ver tabela anexa). Se o bem ou serviço for comprado a um profissional independente, deve ser indicada a sua função, como, por exemplo, médico, babysitter, electricista, empregada de limpeza, massagista, jardineiro, enfermeiro, profissionais de passagem a ferry, advogado, ... No diário geral de consumo do agregado o tipo de estabelecimento não se preenche em caso de recebimentos gratuitos nem de auto-consumo. No diário relativo a saúde preenche-se sempre o estabelecimento.
Se for alimentação (colunas 11 e 12)	Caso o bem adquirido ou recebido seja para alimentação, desde que o seu consumo não ocorra em estabelecimento de restauração, deverá ser assinalado com X o modo de conservação, escolhendo-se uma das 3 opções: natural/fresco ou frigorificado, congelado ou em conserva.

MÓDULO II
DIÁRIO DE CONSUMO DO AGREGADO - SAÚDE
(COMPRAS, RECEBIMENTOS GRATUITOS E AUTO-ABASTECIMENTO)

INSTRUÇÕES DE PREENCHIMENTO GERAIS

Aplicam-se genericamente as mesmas regras descritas para o quadro anterior.

Neste diário deverão ser registados os consumos do agregado em saúde (consumos de natureza frequente, como os medicamentos), anotando-se todos os bens e serviços consumidos, qualquer que seja o seu valor. Se algum indivíduo não quiser preencher o diário individual de saúde, poderão ser aqui incluídos os seus registos (por exemplo, dos indivíduos entre 10 e 14 anos).

INSTRUÇÕES DE PREENCHIMENTO DETALHADAS

Bens ou serviços de saúde obtidos gratuitamente (coluna 3):

É aplicável, por exemplo, a indivíduos isentos de taxa moderadora e ainda a beneficiários de medicina no trabalho.

Valor pago (coluna 8):

Deverá ser registado o valor que efectivamente foi pago pelo indivíduo/agregado.

Se previsivelmente ainda irá ocorrer um reembolso, não se deverá fazer esse abatimento, dado que ainda não se conhece com precisão o valor exacto a receber.

Os bens ou serviços de saúde obtidos gratuitamente, ao contrário de todos os restantes, serão registados a preço zero.

Comparticipação ou preço reduzido? (colunas 10 a 12):

A opção "Não" destina-se aos bens ou serviços que não tiveram nem vão ter qualquer participação do Estado ou outra entidade.

A opção "Sim - já ocorreu" destina-se a bens ou serviços que tiveram logo à partida participação do Estado ou de outra entidade (aplica-se aos medicamentos comparticipados e a todos os serviços médicos prestados pelo Serviço Nacional de Saúde (SNS), incluindo situações de isenção de taxa moderadora).

A opção "Sim - vai ocorrer mais tarde" destina-se a bens ou serviços que, independentemente de já terem tido ou não, à partida, participação do Estado ou de outra entidade, vão ser previsivelmente alvo de reembolso futuro, no todo ou em parte.

Sistema(s) de participação (colunas 13 e 14):

A indicação do(s) sistema(s) deve ser feita tanto para participações que já ocorreram como para situações futuras.

Poderá haver um ou mais sistemas de participação.

Quando há mais do que um sistema de participação, a situação mais frequente é a combinação do Serviço Nacional de Saúde (SNS) com outro sistema. Deste modo, assinala-se X na coluna do SNS e escolhe-se o sistema adicional na tabela anexa.

Havendo 2 ou mais sistemas simultâneos excepto o SNS, deverá ser escolhido o que proporcionar a maior contribuição em valor.

Deverá haver o cuidado de distinguir serviços sociais de seguros. No caso de um instituto público com seguro de saúde, a opção correcta será a 10.

2.1 - DIÁRIO DE CONSUMO DO AGREGADO - EXCEPTO SAÚDE												1ª Semana			Segunda-feira			data/mês / / 00		
EMPRESAS, RECEBIMENTOS GRATUITOS E AUTO-ABASTECIMENTO												Houve registos neste dia? Sim <input type="checkbox"/> Não <input type="checkbox"/>								
EXEMPLOS																				
Nº	Mencionar os produtos consumidos no serviço				Compre ou recebeu grátis? (de 001 até 003)	Descrição do produto ou serviço	Quantidade	Valor (euros)	Tipo de estabelecimento ou de profissional independente (verifique e não altere facilmente)	Se se abastecer (excepto auto-abastecimento)										
	Comida	Bebidas (alcoólicas e não alcoólicas)	Outros produtos	Outros serviços						Não se abastecer	Comparticipa	Outros								
	001	002	003	004					001	002	003									
001	X					Talhetas para computadores para a internet caber-lhes	10 folhas	0,20	2,00	Comércio de equipamentos										
002	X					Talhetas	1	0,20	0,20	Talhetaria										
003	X					Ar condicionado de verão	1	0,00	0,00	ar condicionado										
004		X				Clayton de fundição inglesa	0,2 litros	1,00	0,20		X									
005	X					Talhetas de cinema	2	0,50	1,00	cinema										
006			X			Pavimento de madeira	1	0,00	0,00	de madeira										
007		X				Casa / jardim	2	0,00	0,00											
008	X			X		Imagem - software completo	2 programas	1,00	2,00	software										
009	X					Chaveiro para-lua rápido para ligar 2 mil	1	1,00	1,00	alguma										
010	X					Paqueta de algodão para viagem	1 pac	0,50	0,50	de algodão										
011	X					Limpeza a seco de vestidos	1	1,50	2,00	lavandaria										
012	X					Desodorizante (para o corpo)	1	1,00	0,50	de higiene										
013	X					Reparação de vidros	1	1,00	1,00	de vidro										
014		X				Molho de carne	0,2 kg	1,00	0,20		X									
015	X					Cerveja de casa de 2ª, sem álcool	0,8 kg	1,00	0,80	refrigerado	X									
016	X					Arroz extra longo (comprimento médio)	2 kg	0,50	1,00	-	X									
017	X					Pão de leite de 2ª tipo comum	0,4 kg	1,00	0,40	-	X									
018	X					Folhas de presunto	0,2 kg	1,00	0,20	-		X								
019	X					Alface em lata	0,50 kg	1,00	0,50	-			X							
020	X					Lado de linguiça de casa grande	2 kg	1,00	2,00	-	X									
021	X					Capri salado com cogumelos	0,1 kg	1,00	0,10	-	X									
022	X					Chaves	0,01 kg	1,00	0,01	-	X									
023	X					Talhetas de cinema - Taver	1	0,20	0,20	software										
024	X					Almofa	1	0,50	0,50	software										
025			X			Jornal "comunal"	1	1,00	1,00	de jornais										

Total de visitas com preenchimento: 2 / 15
a preencher por exemplo

2.15 - DIÁRIO DE CONSUMO DO AGREGADO - SAÚDE												1ª Semana			de 11 / 01 a 26 / 00					
EMPRESAS, RECEBIMENTOS GRATUITOS E AUTO-ABASTECIMENTO												Houve registos nesta semana? Sim <input type="checkbox"/> Não <input type="checkbox"/>								
EXEMPLOS																				
Nº	Mencionar os produtos consumidos no dia ou no dia				Compre ou recebeu grátis? (de 001 até 003)	Descrição do produto ou serviço	Quantidade	Valor (euros)	Tipo de estabelecimento ou de profissional independente (verifique e não altere facilmente)	Comparticipação no preço reduzido?										
	Comida	Bebidas (alcoólicas e não alcoólicas)	Outros produtos	Outros serviços						Não	de 001 até 003	de 004 até 006	de 007 até 009							
	001	002	003	004					001	002	003	004	005	006						
001	X					Frango assado	1 unidade	1,00	2,00	restaurante	X									
002	X					Chocolate	1	0,50	0,50	confeitaria	X	X	X	X	20					
003		X				Água engarrafada	1	0,50	0,50	engarrafada	X									
004	X					Frango assado	1 unidade	0,50	0,50	engarrafada	X									
005		X				Molho de tomate	1	0,50	0,50	molho de tomate		X			20					
006	X					Consulta de dermatologia em consultório particular	1	0,00	0,00	medico particular			X		20					
007	X					"Almofa"	1	0,50	0,50	confeitaria		X	X	X	20					
008	X					Consulta de cardiologia	1	0,00	0,00	medico de casa		X			20					
009	X					Consulta com medico particular - almofa grátis	1	0,50	0,00	medico particular		X	X		20					
010																				
011																				
012																				
013																				
014																				
015																				
016																				
017																				
018																				
019																				
020																				

Total de visitas com preenchimento: 2 / 15
a preencher por exemplo

LEGENDA DE SÍMBOLOS DE PARTICIPAÇÃO											
1. ADP (Associação de Defesa dos Direitos do Paciente)	2. Associação de Defesa dos Direitos do Paciente	3. Associação de Defesa dos Direitos do Paciente	4. Associação de Defesa dos Direitos do Paciente	5. Associação de Defesa dos Direitos do Paciente	6. Associação de Defesa dos Direitos do Paciente	7. Associação de Defesa dos Direitos do Paciente	8. Associação de Defesa dos Direitos do Paciente	9. Associação de Defesa dos Direitos do Paciente	10. Associação de Defesa dos Direitos do Paciente	11. Associação de Defesa dos Direitos do Paciente	12. Associação de Defesa dos Direitos do Paciente

2.1 - DIÁRIO DE CONSUMO DO AGREGADO - EXCEPTO SAÚDE (COMPRAS, RECEBIMENTOS GRATUITOS, AUTO-CONSUMO E AUTO-ABASTECIMENTO)										1ª Semana		Segunda-feira		dia/mês ____ / ____		
Houve registos neste dia?										Sim <input type="checkbox"/>		Não <input type="checkbox"/>				
n.º	Modalidade de consumo do bem ou serviço				Compra no estabelecimento? (de não ser caso, indicar N)	Descrição do produto ou serviço	Quantidade	Valor (euros)	Tipo de estabelecimento ou de profissional independente (comprar e auto-abastecimento)	Se for alimentação (descrever a refeição)						
	Compra	Recebeu gratuito	Auto-consumo	Auto-abastecimento						Nacional	Compartilhado	Comensal	Indicação			
													1	2		
001																
002																
003																
004																
005																
006																
007																
008																
009																
010																
011																
012																
013																
014																
015																
016																
017																
018																
019																
020																
021																
022																
023																
024																
025																
Total de linhas com preenchimento																

Total de linhas com preenchimento
apresentar por agregado

2.2 - DIÁRIO DE CONSUMO DO AGREGADO - EXCEPTO SAÚDE (COMPRAS, RECEBIMENTOS GRATUITOS, AUTO-CONSUMO E AUTO-ABASTECIMENTO)										1ª Semana		Terça-feira		dia/mês ____ / ____		
Houve registos neste dia?										Sim <input type="checkbox"/>		Não <input type="checkbox"/>				
n.º	Modalidade de consumo do bem ou serviço				Compra no estabelecimento? (de não ser caso, indicar N)	Descrição do produto ou serviço	Quantidade	Valor (euros)	Tipo de estabelecimento ou de profissional independente (comprar e auto-abastecimento)	Se for alimentação (descrever a refeição)						
	Compra	Recebeu gratuito	Auto-consumo	Auto-abastecimento						Nacional	Compartilhado	Comensal	Indicação			
													1	2		
001																
002																
003																
004																
005																
006																
007																
008																
009																
010																
011																
012																
013																
014																
015																
016																
017																
018																
019																
020																
021																
022																
023																
024																
025																
Total de linhas com preenchimento																

Total de linhas com preenchimento
apresentar por agregado

2.3 - DIÁRIO DE CONSUMO DO AGREGADO - EXCEPTO SAÚDE (COMPRAS, RECEBIMENTOS GRATUITOS, AUTO-CONSUMO E AUTO-ABASTECIMENTO)										1ª Semana			Quarta-feira			dia/mês / ____ / ____		
Houve registos neste dia?										Sim <input type="checkbox"/>			Não <input type="checkbox"/>					
n.º	Modalidade de consumo do bem ou serviço				Compra no estabelecimento? (se não, indicar E)	Descrição do produto ou serviço	Quantidade	Valor (euros)	Tipo de estabelecimento ou de profissional independente (compras e auto-abastecimento)	Se for alimentação (exceto restauração)								
	Compra	Recebeu gratuitamente	Auto-consumo	Auto-abastecimento						Indique o modo de utilização	Normal	Completo	Consumo					
	1	2	3	4	5	6	7	8	9	10	11	12	13					
001																		
002																		
003																		
004																		
005																		
006																		
007																		
008																		
009																		
010																		
011																		
012																		
013																		
014																		
015																		
016																		
017																		
018																		
019																		
020																		
021																		
022																		
023																		
024																		
025																		
Total de linhas com preenchimento																		
a preencher por estabelecimento																		

2.4 - DIÁRIO DE CONSUMO DO AGREGADO - EXCEPTO SAÚDE (COMPRAS, RECEBIMENTOS GRATUITOS, AUTO-CONSUMO E AUTO-ABASTECIMENTO)										1ª Semana			Quinta-feira			dia/mês / ____ / ____		
Houve registos neste dia?										Sim <input type="checkbox"/>			Não <input type="checkbox"/>					
n.º	Modalidade de consumo do bem ou serviço				Compra no estabelecimento? (se não, indicar E)	Descrição do produto ou serviço	Quantidade	Valor (euros)	Tipo de estabelecimento ou de profissional independente (compras e auto-abastecimento)	Se for alimentação (exceto restauração)								
	Compra	Recebeu gratuitamente	Auto-consumo	Auto-abastecimento						Indique o modo de utilização	Normal	Completo	Consumo					
	1	2	3	4	5	6	7	8	9	10	11	12	13					
001																		
002																		
003																		
004																		
005																		
006																		
007																		
008																		
009																		
010																		
011																		
012																		
013																		
014																		
015																		
016																		
017																		
018																		
019																		
020																		
021																		
022																		
023																		
024																		
025																		
Total de linhas com preenchimento																		
a preencher por estabelecimento																		

2.5 - DIÁRIO DE CONSUMO DO AGREGADO - EXCEPTO SAÚDE (COMPRAS, RECEBIMENTOS GRATUITOS, AUTO-CONSUMO E AUTO-ABASTECIMENTOS)										1ª Semana		Sexta-feira		dia/mês ____ / ____	
Houve registros neste dia?										Sim <input type="checkbox"/>		Não <input type="checkbox"/>			
Nº	Modalidade de consumo do bem ou serviço				Compra no estabelecimento? (se não, justificar R)	Descrição do produto ou serviço	Quantidade	Valor (reais)	Tipo de estabelecimento ou de profissional independente (compras e auto-abastecimento)	Se for alimentação (exceto restrição)					
	Compra	Recebu-mento gratuito	Auto-consumo	Auto-abastecimento						Indique o modo de conservação	Recebu-mento gratuito	Compras	Outras		
Indique o modo de conservação															
Indique o modo de conservação															
001															
002															
003															
004															
005															
006															
007															
008															
009															
010															
011															
012															
013															
014															
015															
016															
017															
018															
019															
020															
021															
022															
023															
024															
025															
Total de linhas com preenchimento:															

2.6 - DIÁRIO DE CONSUMO DO AGREGADO - EXCEPTO SAÚDE (COMPRAS, RECEBIMENTOS GRATUITOS, AUTO-CONSUMO E AUTO-ABASTECIMENTOS)										1ª Semana		Sábado		dia/mês ____ / ____	
Houve registros neste dia?										Sim <input type="checkbox"/>		Não <input type="checkbox"/>			
Nº	Modalidade de consumo do bem ou serviço				Compra no estabelecimento? (se não, justificar R)	Descrição do produto ou serviço	Quantidade	Valor (reais)	Tipo de estabelecimento ou de profissional independente (compras e auto-abastecimento)	Se for alimentação (exceto restrição)					
	Compra	Recebu-mento gratuito	Auto-consumo	Auto-abastecimento						Indique o modo de conservação	Recebu-mento gratuito	Compras	Outras		
Indique o modo de conservação															
Indique o modo de conservação															
001															
002															
003															
004															
005															
006															
007															
008															
009															
010															
011															
012															
013															
014															
015															
016															
017															
018															
019															
020															
021															
022															
023															
024															
025															
Total de linhas com preenchimento:															

2.7 - DIÁRIO DE CONSUMO DO AGREGADO - EXCEPTO SAÚDE (COMPRA, RECEBIMENTOS GRATUITOS, AUTO-CONSUMO E AUTO-ABASTECIMENTO)										1ª Semana		Domingo		diária /	
Houve registo neste dia?										Sim <input type="checkbox"/>		Não <input type="checkbox"/>			
Nº	Mozzinhos em consumo de leite ou soroço				Compra no exterior? (se sim, explicar)	Descrição do produto ou serviço	Quantidade	Valor (Mozz)	Tipo de estabelecimento ou de profissional independente (compra e auto-abastecimento)	Se for alimentação (sempre indicar)					
	Compa	Recebi. gratis	Auto consumo	Auto abastecimento						Recebi. gratis ou gratuito	Compartil.	Comensal			
Mozzinhos em consumo de leite ou soroço										Mozzinhos em consumo de leite ou soroço		Mozzinhos em consumo de leite ou soroço		Mozzinhos em consumo de leite ou soroço	
001															
002															
003															
004															
005															
006															
007															
008															
009															
010															
011															
012															
013															
014															
015															
016															
017															
018															
019															
020															
021															
022															
023															
024															
025															
Total de linhas com preenchimento															

2.8 - DIÁRIO DE CONSUMO DO AGREGADO - EXCEPTO SAÚDE (COMPRA, RECEBIMENTOS GRATUITOS, AUTO-CONSUMO E AUTO-ABASTECIMENTO)										2ª Semana		Segunda-feira		diária /	
Houve registo neste dia?										Sim <input type="checkbox"/>		Não <input type="checkbox"/>			
Nº	Mozzinhos em consumo de leite ou soroço				Compra no exterior? (se sim, explicar)	Descrição do produto ou serviço	Quantidade	Valor (Mozz)	Tipo de estabelecimento ou de profissional independente (compra e auto-abastecimento)	Se for alimentação (sempre indicar)					
	Compa	Recebi. gratis	Auto consumo	Auto abastecimento						Recebi. gratis ou gratuito	Compartil.	Comensal			
Mozzinhos em consumo de leite ou soroço										Mozzinhos em consumo de leite ou soroço		Mozzinhos em consumo de leite ou soroço		Mozzinhos em consumo de leite ou soroço	
001															
002															
003															
004															
005															
006															
007															
008															
009															
010															
011															
012															
013															
014															
015															
016															
017															
018															
019															
020															
021															
022															
023															
024															
025															
Total de linhas com preenchimento															

2.9 - DIÁRIO DE CONSUMO DO AGREGADO - EXCEPTO SAÚDE (COMPRAS, RECEBIMENTOS GRATUITOS, AUTO-CONSUMO E AUTO-ABASTECIMENTO)										2ª Semana		Terça-feira		diária ____ / ____		
Houve registos neste dia?										Sim <input type="checkbox"/>		Não <input type="checkbox"/>				
Nº	Localização de consumo do bem ou serviço				Compra no estabelecimento? (se não, especificar E)	Descrição do produto ou serviço	Quantidade	Valor (euros)	Tipo de estabelecimento ou de profissional independente (comprar e auto-abastecimento)	Se for alimentação (excepto medicamentos)						
	Compra	Recebi em gratuidade	Auto-consumo	Auto-abastecimento						Indique o modo de conservação	Recebi em gratuidade	Completar	Consumo			
Indicador I											Indicador II					
001																
002																
003																
004																
005																
006																
007																
008																
009																
010																
011																
012																
013																
014																
015																
016																
017																
018																
019																
020																
021																
022																
023																
024																
025																
Total de linhas com preenchimento																

2.10 - DIÁRIO DE CONSUMO DO AGREGADO - EXCEPTO SAÚDE (COMPRAS, RECEBIMENTOS GRATUITOS, AUTO-CONSUMO E AUTO-ABASTECIMENTO)										2ª Semana		Quarta-feira		diária ____ / ____		
Houve registos neste dia?										Sim <input type="checkbox"/>		Não <input type="checkbox"/>				
Nº	Localização de consumo do bem ou serviço				Compra no estabelecimento? (se não, especificar E)	Descrição do produto ou serviço	Quantidade	Valor (euros)	Tipo de estabelecimento ou de profissional independente (comprar e auto-abastecimento)	Se for alimentação (excepto medicamentos)						
	Compra	Recebi em gratuidade	Auto-consumo	Auto-abastecimento						Indique o modo de conservação	Recebi em gratuidade	Completar	Consumo			
Indicador I											Indicador II					
001																
002																
003																
004																
005																
006																
007																
008																
009																
010																
011																
012																
013																
014																
015																
016																
017																
018																
019																
020																
021																
022																
023																
024																
025																
Total de linhas com preenchimento																

2.11 - DIÁRIO DE CONSUMO DO AGREGADO - EXCEPTO SAÚDE (COMPRAS, RECEBIMENTOS GRATUITOS, AUTO-CONSUMO E AUTO-ABASTECIMENTO)										2ª Semana		Quinta-feira		dia/mês ____ / ____		
Houve registos neste dia?										Sim <input type="checkbox"/>		Não <input type="checkbox"/>				
N.º	Localização do consumo do leite ou serviço				Compra no exterior grátis? (se não, indicar B)	Descrição do produto ou serviço	Quantidade	Valor (euros)	Tipo de estabelecimento ou de profissional independente (comprar e auto-abastecimento)	Se for alimentação (exceto restauração)						
	Compra	Recebi, mas não pagou	Auto-consumo	Auto-abastecimento						Indique o modo de conservação						
										Recebi, mas não pagou	Compartilhado	Comensal				
Indique o dia				Indique o dia			Indique o dia									
001																
002																
003																
004																
005																
006																
007																
008																
009																
010																
011																
012																
013																
014																
015																
016																
017																
018																
019																
020																
021																
022																
023																
024																
025																
Total de linhas com preenchimento																

2.12 - DIÁRIO DE CONSUMO DO AGREGADO - EXCEPTO SAÚDE (COMPRAS, RECEBIMENTOS GRATUITOS, AUTO-CONSUMO E AUTO-ABASTECIMENTO)										2ª Semana		Sexta-feira		dia/mês ____ / ____		
Houve registos neste dia?										Sim <input type="checkbox"/>		Não <input type="checkbox"/>				
N.º	Localização do consumo do leite ou serviço				Compra no exterior grátis? (se não, indicar B)	Descrição do produto ou serviço	Quantidade	Valor (euros)	Tipo de estabelecimento ou de profissional independente (comprar e auto-abastecimento)	Se for alimentação (exceto restauração)						
	Compra	Recebi, mas não pagou	Auto-consumo	Auto-abastecimento						Indique o modo de conservação						
										Recebi, mas não pagou	Compartilhado	Comensal				
Indique o dia				Indique o dia			Indique o dia									
001																
002																
003																
004																
005																
006																
007																
008																
009																
010																
011																
012																
013																
014																
015																
016																
017																
018																
019																
020																
021																
022																
023																
024																
025																
Total de linhas com preenchimento																

2.13 - DIÁRIO DE CONSUMO DO AGREGADO - EXCEPTO SAÚDE
(COMPRAS, RECEBIMENTOS GRATUITOS, AUTO-CONSUMO E AUTO-ABASTECIMENTO)

2ª Semana Sábado dia/mês ____ / ____

Houve registros neste dia? Sim Não

UF	Modalidade de consumo do bem ou serviço				Compras no exterior (em moeda estrangeira R\$)	Descrição do produto ou serviço	Quantidade	Valor (R\$)	Tipo de estabelecimento ou de profissional independente (compras e auto-abastecimento)	Se for alimentação (exceto refeições)			
	Compra	Recebu gratuito	Auto- consumo	Auto- abastecimento						Indique o modo de consumo	Novos francos suíços	Completos	Comensais
Observação													
001													
002													
003													
004													
005													
006													
007													
008													
009													
010													
011													
012													
013													
014													
015													
016													
017													
018													
019													
020													
021													
022													
023													
024													
025													

Tote de linhas com preenchimento:

2.14 - DIÁRIO DE CONSUMO DO AGREGADO - EXCEPTO SAÚDE
(COMPRAS, RECEBIMENTOS GRATUITOS, AUTO-CONSUMO E AUTO-ABASTECIMENTO)

2ª Semana Domingo dia/mês ____ / ____

Houve registros neste dia? Sim Não

UF	Modalidade de consumo do bem ou serviço				Compras no exterior (em moeda estrangeira R\$)	Descrição do produto ou serviço	Quantidade	Valor (R\$)	Tipo de estabelecimento ou de profissional independente (compras e auto-abastecimento)	Se for alimentação (exceto refeições)			
	Compra	Recebu gratuito	Auto- consumo	Auto- abastecimento						Indique o modo de consumo	Novos francos suíços	Completos	Comensais
Observação													
001													
002													
003													
004													
005													
006													
007													
008													
009													
010													
011													
012													
013													
014													
015													
016													
017													
018													
019													
020													
021													
022													
023													
024													
025													

Tote de linhas com preenchimento:

2.15 - DIÁRIO DE CONSUMO DO AGREGADO - SAÚDE (COMPRAS, RECEBIMENTOS GRATUITOS E AUTO-ABASTECIMENTO)						1ª Semana de ____ / ____ a ____ / ____							
Houve registros nesta semana? Sim <input type="checkbox"/> Não <input type="checkbox"/>													
UF	Modalidade de consumo do bem ou serviço			Compra registrada? (de não registrada X)	Descrição do produto ou serviço	Quantidade	Valor pago (R\$)	Tipo de estabelecimento ou de profissional independente (indica-se tipo de consumo)	Participação ou preço reduzido?				
	Compra	Recebido gratuito	Auto-abastecimento						Sim		Sistema(s) de participação		
									Não	Não	Sistema Nacional de Saúde	Outro sistema	
Anexo 100-2				Anexo 2		Anexo 2 (com o anexo 10)		Anexo 2		Anexo 2			
1	2	3	4	5	6	7	8	9	10	11	12	13	14
001													
002													
003													
004													
005													
006													
007													
008													
009													
010													
011													
012													
013													
014													
015													
016													
017													
018													
019													
020													

Total de linhas com preenchimento: _____
de acordo com o formulário

TABELA DE SISTEMAS DE PARTICIPAÇÃO

1. ACSF (Instituição em regime de Terceiro Setor)	4. Serviço Social de Atendimento	7. Serv. Social de empresas para fins sociais (CIP, PT, RDP, CSD, ...)	10. Regime de saúde pago por empregador (serviço médico, Malware, ...)
2. Planos Avulso (OSM, ACMA, ...)	5. Serviço Social Universitário	8. SIA/S e SIA/S Quilombos	11. Medicina de Trabalho
3. Planos Militares (PM, OM, ...)	6. Serviço Social de outras organizações do Poder Judiciário, Ministério Público, ...	9. Regime de saúde pago pelo agregado (serviço médico, Malware, ...)	12. Outros sistemas de participação

2.15 - DIÁRIO DE CONSUMO DO AGREGADO - SAÚDE (COMPRAS, RECEBIMENTOS GRATUITOS E AUTO-ABASTECIMENTO)						2ª Semana de ____ / ____ a ____ / ____							
Houve registros nesta semana? Sim <input type="checkbox"/> Não <input type="checkbox"/>													
UF	Modalidade de consumo do bem ou serviço			Compra registrada? (de não registrada X)	Descrição do produto ou serviço	Quantidade	Valor pago (R\$)	Tipo de estabelecimento ou de profissional independente (indica-se tipo de consumo)	Participação ou preço reduzido?				
	Compra	Recebido gratuito	Auto-abastecimento						Sim		Sistema(s) de participação		
									Não	Não	Sistema Nacional de Saúde	Outro sistema	
Anexo 100-2				Anexo 2		Anexo 2 (com o anexo 10)		Anexo 2		Anexo 2			
1	2	3	4	5	6	7	8	9	10	11	12	13	14
001													
002													
003													
004													
005													
006													
007													
008													
009													
010													
011													
012													
013													
014													
015													
016													
017													
018													
019													
020													

Total de linhas com preenchimento: _____
de acordo com o formulário

TABELA DE SISTEMAS DE PARTICIPAÇÃO

1. ACSF (Instituição em regime de Terceiro Setor)	4. Serviço Social de Atendimento	7. Serv. Social de empresas para fins sociais (CIP, PT, RDP, CSD, ...)	10. Regime de saúde pago por empregador (serviço médico, Malware, ...)
2. Planos Avulso (OSM, ACMA, ...)	5. Serviço Social Universitário	8. SIA/S e SIA/S Quilombos	11. Medicina de Trabalho
3. Planos Militares (PM, OM, ...)	6. Serviço Social de outras organizações do Poder Judiciário, Ministério Público, ...	9. Regime de saúde pago pelo agregado (serviço médico, Malware, ...)	12. Outros sistemas de participação

INQUÉRITO ÀS DESPESAS DAS FAMÍLIAS

- 2005 / 2006 -

MÓDULO III

DIÁRIO DE CONSUMO DO INDIVÍDUO

[AUTO-PREENCHIMENTO DURANTE 1 QUINZENA]

ENQUADRAMENTO DO INQUÉRITO ÀS DESPESAS DAS FAMÍLIAS (IDEF)

O Instituto Nacional de Estatística (INE) realiza inquéritos aos orçamentos familiares desde 1967, com o objectivo de avaliar as condições de vida dos indivíduos residentes em Portugal, bem como conhecer como são aplicados os seus rendimentos em despesas de consumo.

O actual quadro legal (comunitário) estabelece a necessidade de se realizarem estudos periódicos sobre esta realidade, pelo que o INE vem solicitar e agradecer a vossa colaboração enquanto família seleccionada para responder ao inquérito no contexto de um processo de amostragem aleatória.

Salienta-se o seguinte :

Todos os dados individuais recolhidos no âmbito deste inquérito são confidenciais, estando protegidos por segredo estatístico (art.º 5 da Lei nº6/89 de 15 de Abril), e destinando-se apenas à obtenção de resultados globais sobre o consumo dos agregados.

O INE assegura a destruição dos questionários preenchidos, após a obtenção e divulgação de resultados.

O Entrevistador do INE, devidamente credenciado, que periodicamente o visitará, prestará todos os esclarecimentos que necessitar e ajudará a resolver as dúvidas que lhe surjam.

Entrevista	Data de início da quinzena <input type="text"/> / <input type="text"/> / <input type="text"/> 2 0 0 <input type="text"/>		
Entrevistador	Nome <input type="text"/>	Código	<input type="text"/>
Identificação	Área <input type="text"/>	Alojamento <input type="text"/>	Agregado <input type="text"/>
	Nome <input type="text"/>		
	Morada (rua, av, pc, apartado, etc) <input type="text"/>		
	Nº ou lote <input type="text"/> Andar, sala, etc <input type="text"/> Lado <input type="text"/>		
	Localidade <input type="text"/>		
	Código postal <input type="text"/> - <input type="text"/>		
Telefone(s) para contacto <input type="text"/> ou <input type="text"/>			

MÓDULO III

DIÁRIO DE CONSUMO DO INDIVÍDUO - EXCEPTO SAÚDE (COMPRAS, RECEBIMENTOS GRATUITOS E AUTO-ABASTECIMENTO)

INSTRUÇÕES DE PREENCHIMENTO GERAIS

O presente diário de consumo aplica-se a aquisições de bens e serviços em contrapartida de um pagamento (no momento ou posteriormente), a recebimentos gratuitos de bens ou serviços (prendas, refeições, ...) bem como a bens retirados de estabelecimento explorado por algum membro do agregado, desde que não haja pagamento. Estão, portanto, excluídos consumos resultantes de bens provenientes de produção própria, que devem ser anotados no diário do agregado.

Se algum indivíduo, independentemente da idade, não quiser preencher o diário individual, então os seus registos deverão ser efectuados no diário do agregado. Os indivíduos entre 10 e 14 anos poderão também preencher o diário individual mediante autorização dos titulares do poder paternal. Todas as despesas de natureza individual não inscritas em diários individuais terão de estar inscritas no diário do agregado (nomeadamente das crianças com menos de 10 anos).

Todos os bens e serviços consumidos devem ser anotados, qualquer que seja o seu valor.

Deve-se utilizar uma linha para cada bem ou serviço consumido, podendo ser agrupados apenas no caso dos bens em questão serem absolutamente iguais (como por exemplo, 2 maços de tabaco ou 3 bilhetes de cinema).

As anotações neste diário devem ser efectuadas o mais breve possível após cada consumo de um bem ou serviço, de modo a que não sejam esquecidas quaisquer situações, tornando-se assim muito mais fácil obter um preenchimento organizado e exaustivo.

INSTRUÇÕES DE PREENCHIMENTO DETALHADAS

Modalidade do consumo (colunas 2, 3 ou 4)	Uma das três colunas (2, 3 ou 4) necessitam de ser preenchidas com X, consoante se trate de:
Compra (coluna 2)	Aplica-se a aquisições de bens e serviços em contrapartida de um pagamento (no momento ou posteriormente).
Recebimento gratuito (coluna 3)	Aplica-se aos bens e serviços recebidos sem contrapartida de pagamento (prendas, por ex.), e que não sejam um complemento ao salário. É importante não esquecer eventuais refeições recebidas vindas de fora do agregado, quer sejam de família (a viver em outros agregados), de amigos, vizinhos ou instituições de solidariedade.
Auto-abastecimento (coluna 4)	Abrange os bens e serviços provenientes de estabelecimentos explorados por algum membro do agregado e que não sejam pagos (por exemplo, tabaco retirado de loja do agregado, encadernação de livro em tipografia do agregado, ...).
Compra no estrangeiro? (coluna 5)	Esta coluna só deverá ser preenchida se a despesa ocorrer fora de Portugal. Neste caso, preenche-se com X.
Descrição do produto ou serviço (coluna 6)	Deverá ser apresentada uma descrição muito clara, seja qual for a modalidade de consumo, por forma a permitir uma correcta codificação. Resalva-se que só se podem agrupar numa mesma linha produtos absolutamente equivalentes.
Quantidade (coluna 7)	A quantidade deverá ser expressa com a indicação da unidade de medida, sempre que possível de acordo com o explicitado no CUIDOP, como por exemplo, 1 kg, 2 litros, 3 metros, 1 dúzia, 1 pack com 8 unidades, 2 embalagens de 125 ml cada, 1 unidade, No caso dos serviços, a quantidade também deverá ser preenchida, ainda que, na grande maioria das situações, seja "1".
Valor (coluna 8)	O valor deverá ser expresso em Euros, utilizando os dois casos decimais para os cêntimos, e deve referir-se a totalidade da quantidade do bem ou serviço indicados em cada linha. Mesmo que o pagamento não seja efectuado na ocasião, deve ser sempre preenchido o valor total respectivo. Em caso de auto-abastecimento, o valor a preencher será o praticado em caso de venda ou prestação de serviço a um cliente. Em caso de recebimentos gratuitos, o valor a preencher será aquele que se pagaria no estabelecimento mais próximo a que recorreria.
Tipo de estabelecimento ou de profissional independente (coluna 9)	Se for um estabelecimento especializado em determinado ramo, deverá ser descrito desse modo, anotando-se, por exemplo, padaria, feijo, peixaria, pastelaria, restaurante, café, farmácia, tabacaria, papelaria, pronto-a-vestir, sapataria, ... , consultório médico, banco, ginásio, cinema, modista (ver tabela anexa). Se for um estabelecimento não especializado, deverá ser indicado de acordo com a definição que o próprio estabelecimento atribui a ele mesmo, como por exemplo, mercearia, minimercado, super-mercado, hiper-mercado, praça, quiosque de rua, banca em feira, vendedor ambulante (ver tabela anexa). Se o bem ou serviço for comprado a um profissional independente, deve ser indicada a sua função, como, por exemplo, médico, babysitter, electricista, empregado de limpeza, pizzaiolo, jardineiro, enfermeiro, profissional de passeagem a ferro, advogado, O tipo de estabelecimento não se preenche em caso de recebimentos gratuitos.
Se for alimentação (10 a 12)	coluna 10 a 12: Caso o bem adquirido seja para alimentação, desde que o seu consumo não ocorra em estabelecimento de restauração, deverá ser assinalado com X o modo de conservação, escolhendo-se uma das 3 opções: natural/fresco ou frigorificado, congelado ou em conserva.

MÓDULO III
DIÁRIO DE CONSUMO DO INDIVÍDUO - SAÚDE
(COMPRAS, RECEBIMENTOS GRATUITOS E AUTO-ABASTECIMENTO)

INSTRUÇÕES DE PREENCHIMENTO GERAIS

Aplicam-se genericamente as mesmas regras descritas para o quadro anterior.

Neste diário deverão ser registados os consumos do indivíduo com saúde (por exemplo, medicamentos, consultas médicas, serviços de enfermagem, etc.).

Se algum indivíduo, independentemente da idade, não quiser preencher o diário individual de saúde, então os seus registos deverão ser efectuados no diário de saúde do agregado. Os indivíduos entre 10 e 14 anos poderão também preencher o diário individual de saúde mediante autorização dos titulares do poder paternal. Todas as despesas de saúde de natureza individual não inscritas em diários individuais de saúde terão de estar inscritas no diário de saúde do agregado (nomeadamente das crianças com menos de 10 anos).

INSTRUÇÕES DE PREENCHIMENTO DETALHADAS

Bens ou serviços de saúde obtidos gratuitamente (coluna 3):

É aplicável, por exemplo, a indivíduos isentos de taxa moderadora e ainda a beneficiários de medicina no trabalho.

Valor pago (coluna 8):

Deverá ser registado o valor que efectivamente foi pago pelo indivíduo/agregado.

Se previsivelmente ainda irá ocorrer um reembolso, não se deverá fazer esse abatimento, dado que ainda não se conhece com precisão o valor exacto a receber.

Os bens ou serviços de saúde obtidos gratuitamente, ao contrário de todos os restantes, serão registados a preço zero.

Comparticipação ou preço reduzido? (colunas 10 a 12):

A opção "Não" destina-se aos bens ou serviços que não tiveram nem vão ter qualquer participação do Estado ou outra entidade.

A opção "Sim - já ocorreu" destina-se a bens ou serviços que tiveram logo à partida participação do Estado ou de outra entidade (aplica-se aos medicamentos comparticipados e a todos os serviços médicos prestados pelo Serviço Nacional de Saúde (SNS), incluindo situações de isenção de taxa moderadora).

A opção "Sim - vai ocorrer mais tarde" destina-se a bens ou serviços que, independentemente de já terem tido ou não, à partida, participação do Estado ou de outra entidade, vão ser previsivelmente alvo de reembolso futuro, no todo ou em parte.

Sistema(s) de participação (colunas 13 e 14):

A indicação do(s) sistema(s) deve ser feita tanto para participações que já ocorreram como para situações futuras.

Poderá haver um ou mais sistemas de participação.

Quando há mais do que um sistema de participação, a situação mais frequente é a combinação do Serviço Nacional de Saúde (SNS) com outro sistema. Deste modo, assinala-se X na coluna do SNS e escolhe-se o sistema adicional na tabela anexa.

Havendo 2 ou mais sistemas simultâneos excepto o SNS, deverá ser escolhido o que proporcionar a maior contribuição em valor.

Deverá haver o cuidado de distinguir serviços sociais de seguros. No caso de um instituto público com seguro de saúde, a opção correcta será a 10.

3.3. - DIÁRIO DE CONSUMO DO INDIVÍDUO - SAÚDE (EMPRESAS, RECEBIMENTOS GRATUITOS E AUTO-ABASTECIMENTO)										1ª Semana de ___/___ a ___/___				
										Houve registros nesta semana? Sim <input type="checkbox"/> Não <input type="checkbox"/>				
Nº	Recebimento em consumo de bens ou serviços		Compra em estabelecimento? (Se sim, especificar R)	Descrição do produto ou serviço	Quantidade	Valor pago (R\$)	Tipo de estabelecimento ou profissional independente (Indicar o tipo de contribuição)	Participação ou preço reduzido?						
	Compras	Transferência gratuita, mesmo						Outros, através de terceiros	Sim		Sistema de participação			
									Não	Sim	Sim	Outros		
Sim		Não		Sim		Não								
001														
002														
003														
004														
005														
006														
007														
008														
009														
010														
011														
012														
013														
014														
015														
016														
017														
018														
019														
020														
Total de linhas com preenchimento: _____														

TABELA DE SISTEMAS DE PARTICIPAÇÃO

1. CSOP (Contribuição no Consumidor - Benefício do Fornecedor)	4. Serviço Social de Atendimento	7. Sem. Benefício de empresa para parte física (P.F. - RPP - CDD -)	10. Regime de saúde pago por empregado (inclui plano, Medicina -)
2. Projeto Amarelo (CSOP - ACOMPA -)	5. Serviço Social - Operacional	8. IM&M e IM&M Coativos	11. Medicina de Trabalho
3. Projeto Amarelo (PSP - GMR -)	6. Serviço Social de outros segmentos de Plano (Laboração, saúde, previdência -)	9. Regime de saúde pago pelo empregado (inclui plano, Medicina -)	12. Outros sistemas de compartilhamento

3.4. - DIÁRIO DE CONSUMO DO INDIVÍDUO - SAÚDE (EMPRESAS, RECEBIMENTOS GRATUITOS E AUTO-ABASTECIMENTO)										2ª Semana de ___/___ a ___/___				
										Houve registros nesta semana? Sim <input type="checkbox"/> Não <input type="checkbox"/>				
Nº	Recebimento em consumo de bens ou serviços		Compra em estabelecimento? (Se sim, especificar R)	Descrição do produto ou serviço	Quantidade	Valor pago (R\$)	Tipo de estabelecimento ou profissional independente (Indicar o tipo de contribuição)	Participação ou preço reduzido?						
	Compras	Transferência gratuita, mesmo						Outros, através de terceiros	Sim		Sistema de participação			
									Não	Sim	Sim	Outros		
Sim		Não		Sim		Não								
001														
002														
003														
004														
005														
006														
007														
008														
009														
010														
011														
012														
013														
014														
015														
016														
017														
018														
019														
020														
Total de linhas com preenchimento: _____														

TABELA DE SISTEMAS DE PARTICIPAÇÃO

1. CSOP (Contribuição no Consumidor - Benefício do Fornecedor)	4. Serviço Social de Atendimento	7. Sem. Benefício de empresa para parte física (P.F. - RPP - CDD -)	10. Regime de saúde pago por empregado (inclui plano, Medicina -)
2. Projeto Amarelo (CSOP - ACOMPA -)	5. Serviço Social - Operacional	8. IM&M e IM&M Coativos	11. Medicina de Trabalho
3. Projeto Amarelo (PSP - GMR -)	6. Serviço Social de outros segmentos de Plano (Laboração, saúde, previdência -)	9. Regime de saúde pago pelo empregado (inclui plano, Medicina -)	12. Outros sistemas de compartilhamento



INSTITUTO NACIONAL DE ESTATÍSTICA
 PORTUGAL

INQUÉRITO AS DESPESAS DAS FAMÍLIAS

- 2005 / 2006 -

MÓDULO IV

OUTROS CONSUMOS

(MENSAIS, TRIMESTRAIS, ANUAIS, INCLUINDO RECEBIMENTOS GRATUITOS E A TÍTULO DE SALÁRIO)

ENQUADRAMENTO DO INQUÉRITO AS DESPESAS DAS FAMÍLIAS (IDEF)

O Instituto Nacional de Estatística (INE) realiza inquéritos aos orçamentos familiares desde 1967, com o objectivo de avaliar as condições de vida dos indivíduos residentes em Portugal, bem como conhecer como são aplicados os seus rendimentos em despesas de consumo.

O actual quadro legal (comunitário) estabelece a necessidade de se realizarem estudos periódicos sobre esta realidade, pelo que o INE vem solicitar e agradecer a vossa colaboração enquanto família seleccionada para responder ao inquérito no contexto de um processo de amostragem aleatória.

Salienta-se o seguinte :

Todos os dados individuais recolhidos no âmbito deste inquérito são confidenciais, estando protegidos por segredo estatístico (art.º 5 de Lei nº95/89 de 15 de Abril), e destinando-se apenas à obtenção de resultados globais sobre o consumo dos agregados.

O INE assegure a destruição dos questionários preenchidos, após a obtenção e divulgação de resultados.

O Entrevistador do INE, devidamente credenciado, que periodicamente o visitará, prestará todos os esclarecimentos que necessitar e ajudará a resolver as dúvidas que lhe surjam.

Entrevista	Data de início da quinzena: / / 2 0 0						
Entrevistador	Nome _____	Código	_____				
Identificação	Área	_____	Alojamento	_____	Agregado	_____	
	Nome	_____					
	Morada (rua, av, pc, apartado, etc)	_____					
		Nº ou lote	_____	Andar, sala, etc	_____	Lado	_____
	Localidade	_____					
	Código postal	_____					
Telefone(s) para contacto	_____		ou	_____			

MÓDULO IV

OUTROS CONSUMOS - EXCEPTO SAÚDE

(MENSAS, TRIMESTRAIS, ANUAIS, INCLUINDO RECEBIMENTOS GRATUITOS E A TÍTULO DE SALÁRIO)

INSTRUÇÕES DE PREENCHIMENTO GERAIS

Utilizar uma linha para cada bem ou serviço consumido, podendo ser agrupados apenas no caso dos bens em questão serem absolutamente iguais (como por exemplo, 2 meses de conta da electricidade ou pagamento de limpeza de 3 tapetes).

Apesar do fornecimento da informação em questão ser por via de entrevista directa, é recomendável estudar previamente as tabelas anexas sobre os bens e serviços segundo as periodicidades mensal, trimestral e anual.

As ofertas entre membros do agregado familiar não deverão ser registadas.

INSTRUÇÕES DE PREENCHIMENTO DETALHADAS

Modalidade de consumo (colunas 2, 3, 4 e 5)	Uma das quatro colunas (2, 3, 4 ou 5) necessitará de ser preenchida com X, consoante se trate de:	
	Compra de bem ou serviço (coluna 2)	Aplica-se a aquisições de bens e serviços em contrapartida de um pagamento (no momento ou posteriormente).
	Auto-abastecimento (coluna 3)	Abrange os bens e serviços provenientes de estabelecimentos explorados por algum membro do agregado e que não sejam pagos (por exemplo, vestuário retirado de loja do agregado, roupa levada em lavanderia própria, ...).
	Recebido gratuitamente (coluna 4)	Refere-se a bens e serviços recebidos como oferta, de origem externa ao agregado. São ignoradas as ofertas mútuas dentro do próprio agregado.
	Recebido a título de salário ou complemento (coluna 5)	Abrange os bens e serviços que forem recebidos, por algum membro do agregado, como contrapartida de trabalho, seja de natureza regular ou excepcional.
Compra no estrangeiro? (coluna 6)	Esta coluna só deverá ser preenchida se a despesa ocorrer fora de Portugal. Neste caso, preenche-se com X.	
Descrição do produto ou serviço (coluna 7)	Deverá ser apresentada uma descrição muito clara, seja qual for a modalidade de consumo, por forma a permitir uma correcta codificação. Reafirma-se que só se podem agrupar numa mesma linha produtos absolutamente equivalentes.	
Quantidade (coluna 8)	A quantidade deverá ser expressa com a indicação de unidade de medida, sempre que possível de acordo com o explicitado na COICOP como por exemplo, 1 kg, 2 litros, 3 metros, 1 dúzia, 1 pack com 8 unidades, 2 embalagens de 125 ml cada, 1 unidade, No caso dos serviços, a quantidade também deverá ser preenchida, ainda que, na grande maioria das situações, seja "1".	
Valor (coluna 9)	O valor deverá ser expresso em Euros, utilizando as duas casas decimais para os cêntimos, e deve referir-se à totalidade da quantidade dos bens ou serviços indicados em cada linha. Mesmo que o pagamento não seja efectuado na ocasião, deve ser sempre preenchido o valor total respectivo. Em caso de auto-abastecimento, o valor a preencher será o praticado em caso de venda ou prestação de serviço a um cliente. Em caso de bens ou serviços recebidos gratuitamente ou a título de salário ou complemento, o valor a preencher será aquele que o agregado pagaria se tivesse de os adquirir, no estabelecimento mais próximo a que recorria.	
Tipo de estabelecimento ou de profissional independente (coluna 10)	Se for um estabelecimento especializado em determinado ramo, deverá ser descrito desse modo, anotando-se, por exemplo, padaria, talho, peixaria, pastelaria, restaurante, café, farmácia, tabacaria, papelaria, pronto-a-vestir, sapataria, ..., consultório médico, banco, ginásio, cinema, modista (ver tabela anexa). Se for um estabelecimento não especializado, deverá ser indicado de acordo com a definição que o próprio estabelecimento atribui a ele mesmo, como por exemplo, mercearia, mini-mercado, super-mercado, hiper-mercado, praça, quiosque de rua, banco em feira, vendedor ambulante (ver tabela anexa). Se o bem ou serviço for comprado a um profissional independente, deve ser indicada a sua função, como, por exemplo, médico, babysitter, electricista, empregada de limpeza, massagista, jardineiro, enfermeiro, profissional de passagem a ferro, advogado, O tipo de estabelecimento só se preenche em caso de compra ou de auto-abastecimento, ficando excluídas as situações de recebimentos gratuitos ou a título de salário.	
Se for vestuário ou calçado (colunas 11 e 14)	Caso o bem consumido seja um artigo de vestuário ou calçado, deverá ser assinalado com X o utilizador a que se destina, escolhendo-se uma das 4 opções: bebé (menos de 3 anos de idade), criança (3 a 13 anos), homem ou senhora.	

MÓDULO IV

OUTROS CONSUMOS (MENSIS E ANUAIS) - SAÚDE (COMPRAS, RECEBIMENTOS GRATUITOS E AUTO-ABASTECIMENTO)

INSTRUÇÕES DE PREENCHIMENTO GERAIS

Aplicam-se genericamente as regras gerais de preenchimento dos diários de consumo - do agregado e do indivíduo.

INSTRUÇÕES DE PREENCHIMENTO DETALHADAS

Bens ou serviços de saúde obtidos gratuitamente (coluna 3):

É aplicável, por exemplo, a indivíduos isentos de taxa moderadora e ainda a beneficiários de medicina no trabalho.

Valor pago (coluna 8):

Deverá ser registado o valor que efectivamente foi suportado pelo indivíduo/agregado.

Se, após a compra/pagamento, houve recebimento de reembolso, deve ser registado o valor inicialmente pago, antes do reembolso.

Se previsivelmente ainda irá ocorrer um reembolso, não se deverá fazer esse abatimento, dado que ainda não se conhece com precisão o valor exacto a receber.

Os bens ou serviços de saúde obtidos gratuitamente, ao contrário de todos os restantes, serão registados a preço zero.

Comparticipação ou preço reduzido? (colunas 10 a 12):

A opção "Não" destina-se aos bens ou serviços que não tiveram nem vão ter qualquer participação do Estado ou outra entidade.

A opção "Sim - já ocorreu" destina-se a:

- bens ou serviços que tiveram logo à partida participação do Estado ou de outra entidade (aplica-se aos medicamentos comparticipados e a todos os serviços médicos prestados pelo Serviço Nacional de Saúde (SNS), incluindo situações de isenção de taxa moderadora)
- bens ou serviços que, mesmo não tendo tido à partida participação do Estado ou de outra entidade, vieram mais tarde a ser alvo de reembolso.

A opção "Sim - vai ocorrer mais tarde" destina-se a bens ou serviços que, independentemente de já terem tido ou não, à partida, participação do Estado ou de outra entidade, vão ser previsivelmente alvo de reembolso futuro, no todo ou em parte.

Sistema(s) de participação (colunas 13 e 14):

A indicação do(s) sistema(s) deve ser feita tanto para participações que já ocorreram como para situações futuras.

Poderá haver um ou mais sistemas de participação.

Quando há mais do que um sistema de participação, a situação mais frequente é a combinação do Serviço Nacional de Saúde (SNS) com outro sistema. Deste modo, assinala-se X na coluna do SNS e escolhe-se o sistema adicional na tabela anexa.

Havendo 2 ou mais sistemas simultâneos excepto o SNS, deverá ser escolhido o que proporcionar a maior contribuição em valor.

Deverá haver o cuidado de distinguir serviços sociais de seguros. No caso de um instituto público com seguro de saúde, a opção correcta será a 10.

4.3 - CONSUMOS MENSAIS

(INCLUINDO RECEBIMENTOS GRATUITOS E A TÍTULO DE SALÁRIO OU COMPLEMENTO)

Nº	Modalidade de consumo do bem ou serviço				Descrição do produto ou serviço	Quantidade	Valor (R\$)	Tipo de estabelecimento ou profissional independente (empresa e não estabelecimento)
	Compra	Ass. direto (compra)	Resposta gratuita (marca)	Resposta a custo de material ou serviço (marca)				
001								
002								
003								
004								
005								
006								
007								
008								
009								
010								
011								
012								
013								
014								
015								
016								
017								
018								
019								
020								
021								
022								
023								
024								
025								
Total de linhas com preenchimento								

(se houver, por estabelecimento)

4.2 - CONSUMOS TRIMESTRAIS

(INCLUINDO RECEBIMENTOS GRATUITOS E A TÍTULO DE SALÁRIO OU COMPLEMENTO)

Nº	Modalidade de consumo do bem ou serviço				Descrição do produto ou serviço	Quantidade	Valor (R\$)	Tipo de estabelecimento ou profissional independente (empresa e não estabelecimento)	Se for visitado ao campo			
	Compra	Ass. direto (compra)	Resposta gratuita (marca)	Resposta a custo de material ou serviço (marca)					Instituição			
									Tela (R\$)	Cores (R\$)	Tela (R\$)	Cores (R\$)
001												
002												
003												
004												
005												
006												
007												
008												
009												
010												
011												
012												
013												
014												
015												
016												
017												
018												
019												
020												
021												
022												
023												
024												
025												
Total de linhas com preenchimento												

(se houver, por estabelecimento)

4.3- CONSUMOS ANUAIS
(INCLUINDO DESPESAS GRATUITAS E A TÍTULO DE SALÁRIO OU COMPLEMENTO)

Nº	Modalidade de consumo de bens ou serviços				Código do produto ou serviço (ver anexo 3)	Descrição do produto ou serviço	Quantidade	Valor (R\$)	Tipo de estabelecimento ou profissional independente (conforme o anexo 4)
	Compras		Ass. direta (anexo 2)	Ass. indireta (anexo 2)					
	1	2							
001									
002									
003									
004									
005									
006									
007									
008									
009									
010									
011									
012									
013									
014									
015									
016									
017									
018									
019									
020									
021									
022									
023									
024									
025									
Total de linhas com preenchimento									

(apresentar por extenso)

4.4- SAÚDE - CONSUMOS DE PERIODICIDADE MENSAL

Nº	Modalidade de consumo de bens ou serviços				Código do produto ou serviço (ver anexo 3)	Descrição do produto ou serviço	Quantidade	Valor pago (R\$)	Tipo de estabelecimento ou profissional independente (conforme o anexo 4)	Comparticipação ou preço reduzido?			
	Compras		Ass. direta (anexo 2)	Ass. indireta (anexo 2)						Sim	Sim		
	1	2									Sim	Sim	Sim
001													
002													
003													
004													
005													
006													
007													
008													
009													
010													
011													
012													
013													
014													
015													
016													
017													
018													
019													
020													
021													
022													
023													
024													
025													
Total de linhas com preenchimento													

TABELA DE SÍMBOLOS DE PARTICIPAÇÃO

1) Sim - Despesas em nome do beneficiário	2) Sim - Despesas em nome do beneficiário	3) Sim - Despesas em nome do beneficiário	4) Sim - Despesas em nome do beneficiário
5) Sim - Despesas em nome do beneficiário	6) Sim - Despesas em nome do beneficiário	7) Sim - Despesas em nome do beneficiário	8) Sim - Despesas em nome do beneficiário
9) Sim - Despesas em nome do beneficiário	10) Sim - Despesas em nome do beneficiário	11) Sim - Despesas em nome do beneficiário	12) Sim - Despesas em nome do beneficiário
13) Sim - Despesas em nome do beneficiário	14) Sim - Despesas em nome do beneficiário	15) Sim - Despesas em nome do beneficiário	16) Sim - Despesas em nome do beneficiário

4.5 - SAÚDE - CONSUMOS DE PERIODICIDADE ANUAL

UF	Modalidade de consumo do bem ou serviço			Compra ou entrega ganca? (Se sim, indicar o X)	Descrição de produto ou serviço	Quantidade	Valor pago (Total)	Tipo de estabelecimento ou profissional independente (Indicar o tipo de consumo)	Comparticipação ou preço reduzido?				
	Consumo	Resposta gratuita	Outros						Sim		Não		
									Ja ocorreu	Não ocorreu mais	Servico realizado por prazo	Outro sistema	
1	2	3	4	5	6	7	8	9	10	11	12		
001													
002													
003													
004													
005													
006													
007													
008													
009													
010													
011													
012													
013													
014													
015													
016													
017													
018													
019													
020													

Total de linhas com preenchimento: de linhas em evidência

TABELA DE SISTEMAS DE PARTICIPAÇÃO			
1. ACRR (Associação de Consumidores de Saúde)	4. Serviços Sociais de Militares	7. Banc. Saúde de empresas/público/Fundo (FT, FT-ADP, CSD, ...)	10. Regras de saúde pago por empregador (Indicador Milhões -)
2. Preços Atacado (CAMP - ICMS%)	5. Serviços Sociais Universitários	8. SISA e SISA-Qualis	11. Modelo de Trabalho
3. Preços Atacado (PMP - ICM)	6. Serviços Sociais de outros segmentos do Fideis (Indicador em mil reais por usuário -)	9. Regras de saúde pago por usuário (Indicador Milhões -)	12. Outros sistemas de participação

Anexo II

Modelo de Portugal (scripts)

```
> mod1<-glm(desptot~rendtot+npagr+regocup+gurb+conftotal,
family=binomial("logit"))
> summary (mod1)

Call:
glm(formula = desptot ~ rendtot + npagr + regocup + gurb + conftotal,
family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7109  -0.5825   0.1364   0.6172   2.6915

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.985683    0.140556 -35.471 < 2e-16 ***
rendtot      1.712004    0.056405  30.352 < 2e-16 ***
npagr1       0.775953    0.055286  14.035 < 2e-16 ***
regocup2     -0.047144    0.073209  -0.644  0.5196
regocup3     -0.390244    0.075540  -5.166 2.39e-07 ***
regocup4     -0.153272    0.114147  -1.343  0.1793
gurb2        -0.129452    0.070824  -1.828  0.0676 .
gurb3        -0.448064    0.076154  -5.884 4.01e-09 ***
conftotal    0.204882    0.007235  28.318 < 2e-16 ***
---

Null deviance: 14421.6 on 10402 degrees of freedom
Residual deviance: 8903.7 on 10394 degrees of freedom
AIC: 8921.7
```

```
> mod2<-glm(desptot~rmonet+rnaomonet+npagr+regocup+gurb+conftotal,
family=binomial("logit"))
> summary (mod2)

Call:
glm(formula = desptot ~ rmonet + rnaomonet + npagr + regocup + gurb +
conftotal, family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7804  -0.5950   0.1256   0.6198   2.8742

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.611601    0.148642 -37.752 < 2e-16 ***
rmonet1      1.249492    0.057769  21.629 < 2e-16 ***
rnaomonet1   1.628624    0.060991  26.703 < 2e-16 ***
npagr1       0.901052    0.057000  15.808 < 2e-16 ***
regocup2     -0.086237    0.074787  -1.153  0.2489
regocup3     0.151990    0.081364   1.868  0.0618 .
regocup4     -0.129099    0.115415  -1.119  0.2633
gurb2        -0.061644    0.071920  -0.857  0.3914
gurb3        -0.323928    0.078346  -4.135 3.56e-05 ***
conftotal    0.198722    0.007386  26.904 < 2e-16 ***

Null deviance: 14421.6 on 10402 degrees of freedom
Residual deviance: 8595.9 on 10393 degrees of freedom
AIC: 8615.9
```

```

> anova(mod1, mod2, test="Chisq")
Analysis of Deviance Table

Model 1: desptot ~ rendtot + npagr + regocup + gurb + confttotal
Model 2: desptot ~ rmonet + rnaomonet + npagr + regocup + gurb +
confttotal
  Resid. Df Resid. Dev Df Deviance P(>|Chi|)
1     10394     8903.7
2     10393     8595.9  1    307.78 < 2.2e-16 ***

```

```

> mod2i<-
glm(desptot~rmonet+rnaomonet+npagr+gurb2+confttotal+confttotal:npagr+rmo
net:npagr+
+ npagr:regocup+regocup, family=binomial("logit"))
> summary (mod2i)

```

```

Call:
glm(formula = desptot ~ rmonet + rnaomonet + npagr + gurb2 +
confttotal + confttotal:npagr + rmonet:npagr + npagr:regocup +
regocup, family = binomial("logit"))

```

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7517  -0.5882   0.1583   0.6101   2.8486

```

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -6.19869    0.21039  -29.463 < 2e-16 ***
rmonet1        1.41082    0.08709   16.200 < 2e-16 ***
rnaomonet1     1.62902    0.06131   26.570 < 2e-16 ***
npagr1         2.12126    0.28254    7.508 6.01e-14 ***
gurb21        -0.29720    0.07636   -3.892 9.94e-05 ***
confttotal     0.22328    0.01131   19.749 < 2e-16 ***
regocup2       0.16847    0.12553    1.342  0.17958
regocup3       0.23633    0.11485    2.058  0.03962 *
regocup4      -0.11903    0.18296   -0.651  0.51530
npagr1:confttotal -0.04911    0.01493   -3.289  0.00101 **
rmonet1:npagr1 -0.33661    0.11617   -2.897  0.00376 **
npagr1:regocup2 -0.37938    0.15569   -2.437  0.01482 *
npagr1:regocup3 -0.19050    0.14610   -1.304  0.19229
npagr1:regocup4 -0.04911    0.23619   -0.208  0.83529

```

```

Null deviance: 14421.6 on 10402 degrees of freedom
Residual deviance: 8557.3 on 10389 degrees of freedom
AIC: 8585.3

```

```

> mod2ii<-
glm(desptot~rmonet+rnaomonet+npagr+gurb2+confttotal+confttotal:npagr+rmo
net:npagr,
+ family=binomial("logit"))
> summary (mod2ii)

```

```

Call:
glm(formula = desptot ~ rmonet + rnaomonet + npagr + gurb2 +
confttotal + confttotal:npagr + rmonet:npagr, family =
binomial("logit"))

```

```

Deviance Residuals:

```


	Min	1Q	Median	3Q	Max
	-2.7856	-0.5950	0.1576	0.6101	2.8369

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-6.18596	0.20713	-29.865	< 2e-16 ***
rmonet1	1.43139	0.08668	16.512	< 2e-16 ***
rnaomonet1	1.56994	0.05511	28.485	< 2e-16 ***
npagr1	2.19091	0.27707	7.907	2.63e-15 ***
gurb21	-0.31935	0.07488	-4.265	2.00e-05 ***
conftotal	0.22721	0.01109	20.481	< 2e-16 ***
npagr1:conftotal	-0.05824	0.01452	-4.012	6.01e-05 ***
rmonet1:npagr1	-0.36924	0.11553	-3.196	0.00139 **

Null deviance: 14421.6 on 10402 degrees of freedom
Residual deviance: 8570.1 on 10395 degrees of freedom
AIC: 8586.1

```
> anova(mod2ii, mod2i, test="Chisq")
Analysis of Deviance Table

Model 1: desptot ~ rmonet + rnaomonet + npagr + gurb2 + conftotal +
conftotal:npagr +
  rmonet:npagr
Model 2: desptot ~ rmonet + rnaomonet + npagr + gurb2 + conftotal +
conftotal:npagr +
  rmonet:npagr + npagr:regocup + regocup
Resid. Df Resid. Dev Df Deviance P(>|Chi|)
1      10395      8570.1
2      10389      8557.3  6    12.806  0.04622 *
```

```
> modf<-
glm(desptot~rmonet+rnaomonet+npagr+regocup+gurb2+conftotal+conftotal:r
monet+conftotal:rnaomonet+conftotal:npagr+conftotal:regocup+npagr:gurb
2+rmonet:npagr+gurb2:regocup+regocup:rmonet+regocup:rnaomonet+npagr:rn
aomonet+npagr:gurb2,family=binomial("logit"))
> modf<-step (modf)
Start: AIC=8581.71
desptot ~ rmonet + rnaomonet + npagr + regocup + gurb2 + conftotal +
conftotal:rmonet + conftotal:rnaomonet + conftotal:npagr +
conftotal:regocup + npagr:gurb2 + rmonet:npagr + gurb2:regocup +
regocup:rmonet + regocup:rnaomonet + npagr:rnaomonet + npagr:gurb2

Df Deviance AIC
- rnaomonet:npagr 1 8527.7 8579.7
- rmonet:conftotal 1 8528.3 8580.3
- npagr:gurb2 1 8529.5 8581.5
<none> 8527.7 8581.7
- rnaomonet:conftotal 1 8530.1 8582.1
- rnaomonet:regocup 3 8534.3 8582.3
- rmonet:regocup 3 8534.8 8582.8
- regocup:conftotal 3 8535.5 8583.5
- regocup:gurb2 3 8538.5 8586.5
- rmonet:npagr 1 8537.9 8589.9
- npagr:conftotal 1 8540.4 8592.4

Step: AIC=8579.73
desptot ~ rmonet + rnaomonet + npagr + regocup + gurb2 + conftotal +
rmonet:conftotal + rnaomonet:conftotal + npagr:conftotal +
regocup:conftotal + npagr:gurb2 + rmonet:npagr + regocup:gurb2 +
```

```

rmonet:regocup + rnaomonet:regocup

      Df Deviance   AIC
- rmonet:conftotal 1 8528.3 8578.3
- npagr:gurb2      1 8529.5 8579.5
<none>            8527.7 8579.7
- rnaomonet:conftotal 1 8530.1 8580.1
- rnaomonet:regocup 3 8534.4 8580.4
- rmonet:regocup    3 8534.8 8580.8
- regocup:conftotal 3 8535.5 8581.5
- regocup:gurb2     3 8538.5 8584.5
- rmonet:npagr      1 8537.9 8587.9
- npagr:conftotal   1 8540.5 8590.5

Step: AIC=8578.31
desptot ~ rmonet + rnaomonet + npagr + regocup + gurb2 + conftotal +
  rnaomonet:conftotal + npagr:conftotal + regocup:conftotal +
  npagr:gurb2 + rmonet:npagr + regocup:gurb2 + rmonet:regocup +
  rnaomonet:regocup

      Df Deviance   AIC
- npagr:gurb2      1 8530.1 8578.1
<none>            8528.3 8578.3
- rnaomonet:conftotal 1 8530.6 8578.6
- rnaomonet:regocup 3 8535.0 8579.0
- rmonet:regocup    3 8535.7 8579.7
- regocup:conftotal 3 8535.8 8579.8
- regocup:gurb2     3 8539.0 8583.0
- rmonet:npagr      1 8538.6 8586.6
- npagr:conftotal   1 8540.5 8588.5

Step: AIC=8578.11
desptot ~ rmonet + rnaomonet + npagr + regocup + gurb2 + conftotal +
  rnaomonet:conftotal + npagr:conftotal + regocup:conftotal +
  rmonet:npagr + regocup:gurb2 + rmonet:regocup + rnaomonet:regocup

      Df Deviance   AIC
<none>            8530.1 8578.1
- rnaomonet:conftotal 1 8532.4 8578.4
- rnaomonet:regocup 3 8536.9 8578.9
- regocup:conftotal 3 8537.5 8579.5
- rmonet:regocup    3 8537.6 8579.6
- regocup:gurb2     3 8539.8 8581.8
- rmonet:npagr      1 8540.1 8586.1
- npagr:conftotal   1 8541.7 8587.7
> anova(mod, mod2i, test="Chisq")
Analysis of Deviance Table

Model 1: desptot ~ rmonet + rnaomonet + npagr + regocup + gurb2 +
conftotal + rnaomonet:conftotal + npagr:conftotal + regocup:conftotal
+ rmonet:npagr + regocup:gurb2 + rmonet:regocup + rnaomonet:regocup
Model 2: desptot ~ rmonet + rnaomonet + npagr + gurb2 + conftotal +
conftotal:npagr + rmonet:npagr + npagr:regocup + regocup

  Resid. Df Resid. Dev  Df Deviance P(>|Chi|)
1     10379     8530.1
2     10389     8557.3 -10  -27.176  0.002443 **
---
> mod<- glm(desptot~rmonet + rnaomonet + npagr + regocup + conftotal +

```

```

npagr:conftotal + regocup:conftotal + npagr:gurb + rmonet:npagr +
regocup:gurb + rmonet:regocup + rnaomonet:regocup,
family=binomial("logit"))
> summary (mod)

Call:
glm(formula = desptot ~ rmonet + rnaomonet + npagr + regocup +
conftotal + npagr:conftotal + regocup:conftotal + npagr:gurb +
rmonet:npagr + regocup:gurb + rmonet:regocup + rnaomonet:regocup,
family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8661  -0.5820   0.1420   0.6133   2.8760

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -6.50356    0.24550  -26.491 < 2e-16 ***
rmonet1           1.32496    0.09839   13.467 < 2e-16 ***
rnaomonet1       1.59567    0.07548   21.139 < 2e-16 ***
npagr1            2.17582    0.29352    7.413 1.24e-13 ***
regocup2          0.76893    0.44034    1.746 0.080771 .
regocup3          0.66628    0.39340    1.694 0.090331 .
regocup4          0.33978    0.55699    0.610 0.541836
conftotal         0.24754    0.01280   19.343 < 2e-16 ***
npagr1:conftotal -0.05465    0.01488   -3.672 0.000241 ***
regocup2:conftotal -0.04713    0.02033   -2.318 0.020445 *
regocup3:conftotal -0.04678    0.02014   -2.322 0.020221 *
regocup4:conftotal -0.03363    0.02827   -1.190 0.234166
npagr0:gurb2     -0.07971    0.12186   -0.654 0.513044
npagr1:gurb2     -0.12766    0.11393   -1.120 0.262505
npagr0:gurb3     -0.33885    0.12310   -2.753 0.005914 **
npagr1:gurb3     -0.56739    0.12555   -4.519 6.21e-06 ***
rmonet1:npagr1   -0.36660    0.11736   -3.124 0.001786 **
regocup2:gurb2   -0.11773    0.18662   -0.631 0.528130
regocup3:gurb2    0.28274    0.22498    1.257 0.208841
regocup4:gurb2    0.32908    0.28377    1.160 0.246182
regocup2:gurb3    0.17745    0.21129    0.840 0.401014
regocup3:gurb3    0.58910    0.30310    1.944 0.051945 .
regocup4:gurb3    0.96990    0.31155    3.113 0.001851 **
rmonet1:regocup2  0.19892    0.15254    1.304 0.192214
rmonet1:regocup3  0.37845    0.15343    2.467 0.013637 *
rmonet1:regocup4 -0.08966    0.23721   -0.378 0.705449
rnaomonet1:regocup2 -0.02560    0.14604   -0.175 0.860861
rnaomonet1:regocup3  0.83489    0.36826    2.267 0.023383 *
rnaomonet1:regocup4 -0.09882    0.23184   -0.426 0.669929

Null deviance: 14421.6 on 10402 degrees of freedom
Residual deviance: 8525.6 on 10374 degrees of freedom
AIC: 8583.6

```

```

> anova(mod2ii, mod, test="Chisq")
Analysis of Deviance Table

Model 1: desptot ~ rmonet + rnaomonet + npagr + gurb2 + conftotal +
conftotal:npagr +
rmonet:npagr
Model 2: desptot ~ rmonet + rnaomonet + npagr + regocup + conftotal +
npagr:conftotal + regocup:conftotal + npagr:gurb + rmonet:npagr +
regocup:gurb + rmonet:regocup + rnaomonet:regocup
Resid. Df Resid. Dev Df Deviance P(>|Chi|)

```

1	10395	8570.1			
2	10374	8525.6	21	44.441	0.002049 **

```

Modelo Final

> modf<-
glm(desptot~rmonet+rnaomonet+npagr+gurb2+confttotal+confttotal:npagr+rmo
net:npagr, family=binomial("logit"))
> summary (modf)

Call:
glm(formula = desptot ~ rmonet + rnaomonet + npagr + gurb2 +
     confttotal + confttotal:npagr + rmonet:npagr, family =
     binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7856  -0.5950   0.1576   0.6101   2.8369

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -6.18596    0.20713  -29.865 < 2e-16 ***
rmonet1      1.43139    0.08668   16.512 < 2e-16 ***
rnaomonet1   1.56994    0.05511   28.485 < 2e-16 ***
npagr1       2.19091    0.27707    7.907 2.63e-15 ***
gurb21      -0.31935    0.07488   -4.265 2.00e-05 ***
confttotal   0.22721    0.01109   20.481 < 2e-16 ***
npagr1:confttotal -0.05824    0.01452   -4.012 6.01e-05 ***
rmonet1:npagr1 -0.36924    0.11553   -3.196 0.00139 **
---

Null deviance: 14421.6 on 10402 degrees of freedom
Residual deviance: 8570.1 on 10395 degrees of freedom
AIC: 8586.1

```

```

Matriz de covariância

> vcov(modf)
              (Intercept)          rmonet1          rnaomonet1
npagr1
(Intercept)  0.042904452  0.0020694546 -1.467503e-03 -
0.0420191622
rmonet1      0.002069455  0.0075142794  4.119100e-04 -
0.0022044777
rnaomonet1   -0.001467503  0.0004119100  3.037560e-03
0.0006543783
npagr1       -0.042019162 -0.0022044777  6.543783e-04
0.0767693436
gurb21       -0.001939874  0.0001002255  2.421435e-05
0.0005013967
confttotal   -0.002189843 -0.0002823819 -1.959237e-05
0.0021807617
npagr1:confttotal 0.002197594  0.0002784318 -1.239204e-05 -
0.0038499008
rmonet1:npagr1 -0.002268617 -0.0074694029 -9.871809e-05
0.0011003221
              gurb21          confttotal npagr1:confttotal
rmonet1:npagr1
(Intercept)  -1.939874e-03 -2.189843e-03  2.197594e-03 -
2.268617e-03

```

```

rmonet1      1.002255e-04 -2.823819e-04      2.784318e-04 -
7.469403e-03
rnaomonet1   2.421435e-05 -1.959237e-05      -1.239204e-05 -
9.871809e-05
npagr1       5.013967e-04  2.180762e-03      -3.849901e-03
1.100322e-03
gurb21       5.607091e-03  5.593839e-05      -3.379771e-05
4.142402e-05
conftotal    5.593839e-05  1.230670e-04      -1.226366e-04
2.817539e-04
npagr1:conftotal -3.379771e-05 -1.226366e-04      2.106993e-04 -
4.085296e-04
rmonet1:npagr1 4.142402e-05  2.817539e-04      -4.085296e-04
1.334784e-02
> a<-summary (modf)
> modf$coef-1.96*sqrt(diag(a$cov.scaled))
      (Intercept)      rmonet1      rnaomonet1      npagr1
-6.59194409      1.26148259      1.46191877      1.64784615
      gurb21      conftotal npagr1:conftotal rmonet1:npagr1
-0.46611580      0.20546871      -0.08669205      -0.59568531
> modf$coef+1.96*sqrt(diag(a$cov.scaled))
      (Intercept)      rmonet1      rnaomonet1      npagr1
-6.59194409      1.26148259      1.46191877      1.64784615
      gurb21      conftotal npagr1:conftotal rmonet1:npagr1
-0.46611580      0.20546871      -0.08669205      -0.59568531
> exp(modf$coef-1.96*sqrt(diag(a$cov.scaled)))
      (Intercept)      rmonet1      rnaomonet1      npagr1
0.001371371      3.530652125      4.314229622      5.195776832
      gurb21      conftotal npagr1:conftotal rmonet1:npagr1
0.627434626      1.228100555      0.916959432      0.551184705
> exp(modf$coef+1.96*sqrt(diag(a$cov.scaled)))
      (Intercept)      rmonet1      rnaomonet1      npagr1
0.003088780      4.959413905      5.354653049      15.393903803
      gurb21      conftotal npagr1:conftotal rmonet1:npagr1
0.841487451      1.282684884      0.970648038      0.866930414

```

```

> tp<-sum(desptot==1)
> tp
[1] 5202
> vp<-sum((modf$fitted>=0.52) & (desptot==1))
> vp
[1] 4214
> vp<-sum((modf$fitted>=0.55) & (desptot==1))
> vp
[1] 4090
> vp<-sum((modf$fitted>=0.60) & (desptot==1))
> vp
[1] 3918
> vp<-sum((modf$fitted>=0.45) & (desptot==1))
> vp
[1] 4381
> vp<-sum((modf$fitted>=0.40) & (desptot==1))
> vp
[1] 4562
> sensb<-vp/tp
> sensb*100
[1] 87.69704
> tn<-sum(desptot==0)
> tn
[1] 5201
> vn<-sum((modf$fitted<0.52) & (desptot==0))

```

```

> vn
[1] 4176
> vn<-sum((modf$fitted<0.55) & (desptot==0))
> vn
[1] 4279
> vn<-sum((modf$fitted<0.60) & (desptot==0))
> vn
[1] 4426
> vn<-sum((modf$fitted<0.45) & (desptot==0))
> vn
[1] 4006
> vn<-sum((modf$fitted<0.40) & (desptot==0))
> vn
[1] 3799
>
> esp<-vn/tn
> esp*100
[1] 73.04365

```

Linearização da variável contínua

```

> summary (conftotal)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.00  15.00  19.00  19.02  23.00  35.00
> conftotal_q<-factor(conftotal_q)

> modf<- glm(desptot~rmonet+rnaomonet+npagr+gurb2+conftotal_q,
family=binomial("logit"))
> summary (modf)

Call:
glm(formula = desptot ~ rmonet + rnaomonet + npagr + gurb2 +
  conftotal_q, family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4983  -0.6060   0.3004   0.6437   2.6776

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.20821    0.08972  -35.759 < 2e-16 ***
rmonet1      1.31256    0.05725   22.926 < 2e-16 ***
rnaomonet1   1.60668    0.05467   29.390 < 2e-16 ***
npagr1       0.90062    0.05595   16.097 < 2e-16 ***
gurb21      -0.34844    0.07425   -4.693 2.7e-06 ***
conftotal_q1 0.91749    0.09382    9.779 < 2e-16 ***
conftotal_q2 1.64467    0.09385   17.524 < 2e-16 ***
conftotal_q3 2.46397    0.10264   24.007 < 2e-16 ***

Null deviance: 14421.6 on 10402 degrees of freedom
Residual deviance: 8707.4 on 10395 degrees of freedom
AIC: 8723.4

```

Bondade de ajustamento

```

> hosmerlem(desptot, fitted(modf))
      X^2      Df    P(>Chi)
2.7609456 8.0000000 0.9484524
>

```

Anexo III

Modelo de Espanha (scripts)

```
> modi<-
glm(esdesptot~esrendtot+esnpagr+esregocup+esgurb+esrendtot:esnpagr,
family=binomial("logit"))
> summary (modi)
Call:
glm(formula = esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
  esrendtot:esnpagr, family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9715  -0.6785   0.5560   0.7660   2.3932

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.53310    0.04266  -35.940 < 2e-16 ***
esrendtot1         2.07662    0.05668   36.640 < 2e-16 ***
esnpagr1           1.41657    0.05041   28.103 < 2e-16 ***
esregocup2         0.29372    0.04042    7.266 3.70e-13 ***
esregocup3        -0.85259    0.06014  -14.176 < 2e-16 ***
esregocup4        -0.16147    0.07867   -2.052  0.0401 *
esgurb2           -0.11234    0.04472   -2.512  0.0120 *
esgurb3           -0.41934    0.04048  -10.359 < 2e-16 ***
esrendtot1:esnpagr1 -0.46487    0.07211   -6.447 1.14e-10 ***
---

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 26943  on 19434  degrees of freedom
Residual deviance: 20304  on 19426  degrees of freedom
AIC: 20322
```

```
> modii<-
glm(esdesptot~esrendtot+esnpagr+esregocup+esgurb+esrendtot:esregocup,
family=binomial("logit"))
> summary (modii)

Call:
glm(formula = esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
  esrendtot:esregocup, family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0007  -0.7096   0.5389   0.7426   2.3712

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.42814    0.04005  -35.663 < 2e-16 ***
esrendtot1         1.81344    0.04714   38.472 < 2e-16 ***
esnpagr1           1.18769    0.03613   32.873 < 2e-16 ***
esregocup2         0.38069    0.06004    6.341 2.29e-10 ***
esregocup3        -0.89542    0.08966   -9.987 < 2e-16 ***
esregocup4        -0.12371    0.10057   -1.230  0.2186
esgurb2           -0.11484    0.04468   -2.570  0.0102 *
esgurb3           -0.42571    0.04049  -10.515 < 2e-16 ***
esrendtot1:esregocup2 -0.09745    0.08081   -1.206  0.2278
esrendtot1:esregocup3  0.12166    0.12191    0.998  0.3183
esrendtot1:esregocup4 -0.05002    0.16018   -0.312  0.7548
(Dispersion parameter for binomial family taken to be 1)
```

```

Null deviance: 26943 on 19434 degrees of freedom
Residual deviance: 20343 on 19424 degrees of freedom
AIC: 20365

```

```

> modiii<-
glm(esdesptot~esrendtot+esnpagr+esregocup+esgurb+esrendtot:esgurb,
family=binomial("logit"))
> summary (modiii)

Call:
glm(formula = esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
  esrendtot:esgurb, family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0192  -0.7195   0.5282   0.7520   2.3344

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.433469   0.043448  -32.993 < 2e-16 ***
esrendtot1         1.815239   0.051793   35.048 < 2e-16 ***
esnpagr1           1.192602   0.035817   33.297 < 2e-16 ***
esregocup2         0.324815   0.040178    8.084 6.25e-16 ***
esregocup3        -0.827442   0.060302  -13.722 < 2e-16 ***
esregocup4        -0.144589   0.078442   -1.843  0.0653 .
esgurb2           -0.110623   0.063656   -1.738  0.0822 .
esgurb3           -0.396103   0.056246   -7.042 1.89e-12 ***
esrendtot1:esgurb2 -0.006281   0.089288   -0.070  0.9439
esrendtot1:esgurb3 -0.059688   0.079884   -0.747  0.4550
---
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 26943 on 19434 degrees of freedom
Residual deviance: 20345 on 19425 degrees of freedom
AIC: 20365

```

```

glm(formula = esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
  esnpagr:esregocup, family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9658  -0.7168   0.5594   0.7248   2.2874

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.50567   0.04149  -36.293 < 2e-16 ***
esrendtot1         1.77099   0.03543   49.981 < 2e-16 ***
esnpagr1           1.36102   0.04759   28.600 < 2e-16 ***
esregocup2         0.60601   0.06467    9.370 < 2e-16 ***
esregocup3        -0.61085   0.09631   -6.343 2.26e-10 ***
esregocup4        -0.09133   0.12569   -0.727  0.46749
esgurb2           -0.11134   0.04470   -2.491  0.01275 *
esgurb3           -0.42362   0.04055  -10.447 < 2e-16 ***
esnpagr1:esregocup2 -0.45654   0.08190   -5.574 2.49e-08 ***
esnpagr1:esregocup3 -0.36094   0.12232   -2.951  0.00317 **
esnpagr1:esregocup4 -0.10697   0.16116   -0.664  0.50687
---
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 26943 on 19434 degrees of freedom

```



```
Residual deviance: 20312 on 19424 degrees of freedom
AIC: 20334
```

```
> modiaa<-
glm(esdesptot~esrendtot+esnpagr+esregocup+esgurb+esnpagr:esgurb,
family=binomial("logit"))
> summary (modiaa)

Call:
glm(formula = esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
    esnpagr:esgurb, family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0040  -0.7140   0.5369   0.7394   2.3550

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.39311    0.04305  -32.362 < 2e-16 ***
esrendtot1    1.79357    0.03524   50.889 < 2e-16 ***
esnpagr1      1.13925    0.05140   22.165 < 2e-16 ***
esregocup2    0.32424    0.04016    8.073 6.85e-16 ***
esregocup3   -0.82853    0.06021  -13.762 < 2e-16 ***
esregocup4   -0.14528    0.07852   -1.850  0.0643 .
esgurb2      -0.16793    0.06941   -2.419  0.0156 *
esgurb3      -0.48695    0.06431   -7.572 3.68e-14 ***
esnpagr1:esgurb2 0.09415    0.09067    1.038  0.2991
esnpagr1:esgurb3 0.10452    0.08222    1.271  0.2036
---

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 26943 on 19434 degrees of freedom
Residual deviance: 20344 on 19425 degrees of freedom
AIC: 20364
```

```
> modib<-
glm(esdesptot~esrendtot+esnpagr+esregocup+esgurb+esregocup:esgurb,
family=binomial("logit"))
> summary (modib)

Call:
glm(formula = esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
    esregocup:esgurb, family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0031  -0.7178   0.5375   0.7707   2.3179

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.33910    0.04177  -32.058 < 2e-16 ***
esrendtot1    1.79787    0.03528   50.960 < 2e-16 ***
esnpagr1      1.18922    0.03584   33.177 < 2e-16 ***
esregocup2    0.11450    0.05807    1.972 0.048649 *
esregocup3   -0.94507    0.08176  -11.559 < 2e-16 ***
esregocup4   -0.24406    0.12620   -1.934 0.053119 .
esgurb2      -0.23360    0.06071   -3.848 0.000119 ***
esgurb3      -0.58624    0.05266  -11.132 < 2e-16 ***
esregocup2:esgurb2 0.33285    0.10123    3.288 0.001009 **
```

```

esregocup3:esgurb2  0.18252    0.15076    1.211  0.226041
esregocup4:esgurb2 -0.03036    0.20785   -0.146  0.883856
esregocup2:esgurb3  0.44416    0.09349    4.751  2.03e-06 ***
esregocup3:esgurb3  0.25477    0.14966    1.702  0.088686 .
esregocup4:esgurb3  0.29167    0.17797    1.639  0.101231
---

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 26943 on 19434 degrees of freedom
Residual deviance: 20318 on 19421 degrees of freedom
AIC: 20346

```

```

> modil<-
glm(esdesptot~esrendtot+esnpagr+esregocup+esgurb+esrendtot:esnpagr+esn
pagr:esregocup, family=binomial("logit"))
> summary (modil)
Call:
glm(formula = esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
esrendtot:esnpagr + esnpagr:esregocup, family = binomial("logit"))

```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.9371	-0.7382	0.5766	0.7463	2.3376

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.59430	0.04527	-35.216	< 2e-16 ***
esrendtot1	2.02147	0.05754	35.133	< 2e-16 ***
esnpagr1	1.53567	0.05725	26.823	< 2e-16 ***
esregocup2	0.54343	0.06725	8.081	6.45e-16 ***
esregocup3	-0.65173	0.09861	-6.609	3.86e-11 ***
esregocup4	-0.08648	0.12842	-0.673	0.50071
esgurb2	-0.11010	0.04474	-2.461	0.01386 *
esgurb3	-0.41893	0.04055	-10.331	< 2e-16 ***
esrendtot1:esnpagr1	-0.40786	0.07288	-5.596	2.19e-08 ***
esnpagr1:esregocup2	-0.38843	0.08366	-4.643	3.43e-06 ***
esnpagr1:esregocup3	-0.31952	0.12348	-2.588	0.00966 **
esnpagr1:esregocup4	-0.13211	0.16243	-0.813	0.41604

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 26943 on 19434 degrees of freedom
Residual deviance: 20280 on 19423 degrees of freedom
AIC: 20304

```

```

> modi2<-
glm(esdesptot~esrendtot+esnpagr+esregocup+esgurb+esrendtot:esnpagr,
family=binomial("logit"))
> summary (modi2)

```

Call:

```

glm(formula = esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
esrendtot:esnpagr, family = binomial("logit"))

```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.9715	-0.6785	0.5560	0.7660	2.3932

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.53310    0.04266 -35.940 < 2e-16 ***
esrendtot1     2.07662    0.05668  36.640 < 2e-16 ***
esnpagr1       1.41657    0.05041  28.103 < 2e-16 ***
esregocup2     0.29372    0.04042   7.266 3.70e-13 ***
esregocup3    -0.85259    0.06014 -14.176 < 2e-16 ***
esregocup4    -0.16147    0.07867  -2.052  0.0401 *
esgurb2       -0.11234    0.04472  -2.512  0.0120 *
esgurb3       -0.41934    0.04048 -10.359 < 2e-16 ***
esrendtot1:esnpagr1 -0.46487    0.07211  -6.447 1.14e-10 ***
---

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 26943 on 19434 degrees of freedom
Residual deviance: 20304 on 19426 degrees of freedom
AIC: 20322

```

```

> anova(modi1, modi2, test="Chisq")
Analysis of Deviance Table

Model 1: esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
esrendtot:esnpagr + esnpagr:esregocup
Model 2: esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
esrendtot:esnpagr
  Resid. Df Resid. Dev Df Deviance P(>|Chi|)
1      19423      20280
2      19426      20304 -3  -23.877  2.65e-05 ***

```

```

> modi<-
glm(esdesptot~esrendtot+esnpagr+esregocup+esgurb+esrendtot:esnpagr,
family=binomial("logit"))
> summary (modi)
Call:
glm(formula = esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
esrendtot:esnpagr, family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9715  -0.6785   0.5560   0.7660   2.3932

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.53310    0.04266 -35.940 < 2e-16 ***
esrendtot1     2.07662    0.05668  36.640 < 2e-16 ***
esnpagr1       1.41657    0.05041  28.103 < 2e-16 ***
esregocup2     0.29372    0.04042   7.266 3.70e-13 ***
esregocup3    -0.85259    0.06014 -14.176 < 2e-16 ***
esregocup4    -0.16147    0.07867  -2.052  0.0401 *
esgurb2       -0.11234    0.04472  -2.512  0.0120 *
esgurb3       -0.41934    0.04048 -10.359 < 2e-16 ***
esrendtot1:esnpagr1 -0.46487    0.07211  -6.447 1.14e-10 ***
---

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 26943 on 19434 degrees of freedom
Residual deviance: 20304 on 19426 degrees of freedom

```

```
AIC: 20322
```

```
> modia<-
glm(esdesptot~esrendtot+esnpagr+esregocup+esgurb+esnpagr:esregocup,
family=binomial("logit"))
> summary (modia)

Call:
glm(formula = esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
     esnpagr:esregocup, family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9658  -0.7168   0.5594   0.7248   2.2874

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.50567    0.04149  -36.293 < 2e-16 ***
esrendtot1         1.77099    0.03543   49.981 < 2e-16 ***
esnpagr1           1.36102    0.04759   28.600 < 2e-16 ***
esregocup2         0.60601    0.06467    9.370 < 2e-16 ***
esregocup3        -0.61085    0.09631   -6.343 2.26e-10 ***
esregocup4        -0.09133    0.12569   -0.727  0.46749
esgurb2           -0.11134    0.04470   -2.491  0.01275 *
esgurb3           -0.42362    0.04055  -10.447 < 2e-16 ***
esnpagr1:esregocup2 -0.45654    0.08190   -5.574 2.49e-08 ***
esnpagr1:esregocup3 -0.36094    0.12232   -2.951  0.00317 **
esnpagr1:esregocup4 -0.10697    0.16116   -0.664  0.50687
---

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 26943  on 19434  degrees of freedom
Residual deviance: 20312  on 19424  degrees of freedom
AIC: 20334
```

```
> modib<-
glm(esdesptot~esrendtot+esnpagr+esregocup+esgurb+esregocup:esgurb,
family=binomial("logit"))
> summary (modib)

Call:
glm(formula = esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
     esregocup:esgurb, family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0031  -0.7178   0.5375   0.7707   2.3179

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.33910    0.04177  -32.058 < 2e-16 ***
esrendtot1         1.79787    0.03528   50.960 < 2e-16 ***
esnpagr1           1.18922    0.03584   33.177 < 2e-16 ***
esregocup2         0.11450    0.05807    1.972 0.048649 *
esregocup3        -0.94507    0.08176  -11.559 < 2e-16 ***
esregocup4        -0.24406    0.12620   -1.934 0.053119 .
esgurb2           -0.23360    0.06071   -3.848 0.000119 ***
esgurb3           -0.58624    0.05266  -11.132 < 2e-16 ***
esregocup2:esgurb2  0.33285    0.10123    3.288 0.001009 **
```

```

esregocup3:esgurb2  0.18252    0.15076    1.211  0.226041
esregocup4:esgurb2 -0.03036    0.20785   -0.146  0.883856
esregocup2:esgurb3  0.44416    0.09349    4.751  2.03e-06 ***
esregocup3:esgurb3  0.25477    0.14966    1.702  0.088686 .
esregocup4:esgurb3  0.29167    0.17797    1.639  0.101231

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 26943 on 19434 degrees of freedom
Residual deviance: 20318 on 19421 degrees of freedom
AIC: 20346

```

```

> moda<-
glm(esdesptot~esrendtot+esnpagr+esregocup+esgurb+esrendtot:esnpagr+esn
pagr:esregocup, family=binomial("logit"))
> summary (moda)

```

Call:

```

glm(formula = esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
esrendtot:esnpagr + esnpagr:esregocup, family = binomial("logit"))

```

Deviance Residuals:

```

      Min       1Q   Median       3Q      Max
-1.9371  -0.7382   0.5766   0.7463   2.3376

```

Coefficients:

```

              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.59430    0.04527  -35.216 < 2e-16 ***
esrendtot1     2.02147    0.05754   35.133 < 2e-16 ***
esnpagr1       1.53567    0.05725   26.823 < 2e-16 ***
esregocup2     0.54343    0.06725    8.081 6.45e-16 ***
esregocup3    -0.65173    0.09861   -6.609 3.86e-11 ***
esregocup4    -0.08648    0.12842   -0.673  0.50071
esgurb2       -0.11010    0.04474   -2.461  0.01386 *
esgurb3       -0.41893    0.04055  -10.331 < 2e-16 ***
esrendtot1:esnpagr1 -0.40786    0.07288   -5.596 2.19e-08 ***
esnpagr1:esregocup2 -0.38843    0.08366   -4.643 3.43e-06 ***
esnpagr1:esregocup3 -0.31952    0.12348   -2.588  0.00966 **
esnpagr1:esregocup4 -0.13211    0.16243   -0.813  0.41604

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 26943 on 19434 degrees of freedom
Residual deviance: 20280 on 19423 degrees of freedom
AIC: 20304

```

```

> modb<-
glm(esdesptot~esrendtot+esnpagr+esregocup+esgurb+esnpagr:esregocup+esr
egocup:esgurb, family=binomial("logit"))
> summary (modb)

```

Call:

```

glm(formula = esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
esnpagr:esregocup + esregocup:esgurb, family = binomial("logit"))

```

Deviance Residuals:

```

      Min       1Q   Median       3Q      Max
-1.9517  -0.7369   0.5678   0.7517   2.2520

```

Coefficients:

```

                Estimate Std. Error z value Pr(>|z|)
(Intercept)    -1.41895    0.04435 -31.993 < 2e-16 ***
esrendtot1     1.77109    0.03549  49.901 < 2e-16 ***
esnpagr1       1.38187    0.04810  28.728 < 2e-16 ***
esregocup2     0.40053    0.07421   5.397 6.77e-08 ***
esregocup3    -0.73234    0.10669  -6.864 6.69e-12 ***
esregocup4    -0.17907    0.15463  -1.158 0.24686
esgurb2       -0.25226    0.06162  -4.094 4.24e-05 ***
esgurb3       -0.61452    0.05365 -11.455 < 2e-16 ***
esnpagr1:esregocup2 -0.51144    0.08232  -6.213 5.21e-10 ***
esnpagr1:esregocup3 -0.39562    0.12329  -3.209 0.00133 **
esnpagr1:esregocup4 -0.14734    0.16188  -0.910 0.36272
esregocup2:esgurb2 0.37263    0.10086   3.695 0.00022 ***
esregocup3:esgurb2 0.22018    0.14999   1.468 0.14211
esregocup4:esgurb2 -0.01176    0.20827  -0.056 0.95499
esregocup2:esgurb3 0.50755    0.09346   5.431 5.61e-08 ***
esregocup3:esgurb3 0.31263    0.14939   2.093 0.03637 *
esregocup4:esgurb3 0.31358    0.17938   1.748 0.08045 .
---

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 26943 on 19434 degrees of freedom
Residual deviance: 20276 on 19418 degrees of freedom
AIC: 20310

```

```

> modc<-
glm(esdesptot~esrendtot+esnpagr+esregocup+esgurb+esrendtot:esnpagr+esr
egocup:esgurb, family=binomial("logit"))
> summary (modc)

Call:
glm(formula = esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
     esrendtot:esnpagr + esregocup:esgurb, family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9616  -0.6746   0.5619   0.7915   2.3681

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    -1.44923    0.04607 -31.457 < 2e-16 ***
esrendtot1     2.07762    0.05673  36.621 < 2e-16 ***
esnpagr1       1.41306    0.05052  27.971 < 2e-16 ***
esregocup2     0.08463    0.05836   1.450 0.146994
esregocup3    -0.97024    0.08174 -11.869 < 2e-16 ***
esregocup4    -0.25925    0.12687  -2.043 0.041015 *
esgurb2       -0.23277    0.06099  -3.816 0.000135 ***
esgurb3       -0.57961    0.05285 -10.966 < 2e-16 ***
esrendtot1:esnpagr1 -0.46173    0.07220  -6.395 1.60e-10 ***
esregocup2:esgurb2 0.33439    0.10126   3.302 0.000959 ***
esregocup3:esgurb2 0.18110    0.15008   1.207 0.227554
esregocup4:esgurb2 -0.03300    0.20837  -0.158 0.874180
esregocup2:esgurb3 0.43494    0.09335   4.659 3.17e-06 ***
esregocup3:esgurb3 0.25750    0.14876   1.731 0.083451 .
esregocup4:esgurb3 0.28740    0.17835   1.611 0.107092
---

(Dispersion parameter for binomial family taken to be 1)

```

```

Null deviance: 26943 on 19434 degrees of freedom
Residual deviance: 20276 on 19420 degrees of freedom
AIC: 20306

```

```

> modabc<-
glm(esdesptot~esrendtot+esnpagr+esregocup+esgurb+esrendtot:esnpagr+esn
pagr:esregocup+esnpagr:esregocup+esregocup:esgurb+esrendtot:esnpagr+es
regocup:esgurb, family=binomial("logit"))
> summary(modabc)

```

```

Call:
glm(formula = esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
esrendtot:esnpagr + esnpagr:esregocup + esnpagr:esregocup +
esregocup:esgurb + esrendtot:esnpagr + esregocup:esgurb,
family = binomial("logit"))

```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.9251	-0.7379	0.5839	0.7723	2.3009

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.50674	0.04804	-31.364	< 2e-16	***
esrendtot1	2.01423	0.05761	34.965	< 2e-16	***
esnpagr1	1.55077	0.05762	26.912	< 2e-16	***
esregocup2	0.34498	0.07632	4.520	6.18e-06	***
esregocup3	-0.76946	0.10868	-7.080	1.44e-12	***
esregocup4	-0.17319	0.15683	-1.104	0.269465	
esgurb2	-0.24910	0.06175	-4.034	5.49e-05	***
esgurb3	-0.60502	0.05371	-11.264	< 2e-16	***
esrendtot1:esnpagr1	-0.39600	0.07300	-5.425	5.80e-08	***
esnpagr1:esregocup2	-0.44349	0.08410	-5.273	1.34e-07	***
esnpagr1:esregocup3	-0.35453	0.12440	-2.850	0.004373	**
esnpagr1:esregocup4	-0.17168	0.16306	-1.053	0.292395	
esregocup2:esgurb2	0.36796	0.10098	3.644	0.000268	***
esregocup3:esgurb2	0.21429	0.14966	1.432	0.152189	
esregocup4:esgurb2	-0.01495	0.20814	-0.072	0.942722	
esregocup2:esgurb3	0.48994	0.09348	5.241	1.59e-07	***
esregocup3:esgurb3	0.30764	0.14887	2.067	0.038780	*
esregocup4:esgurb3	0.31198	0.17918	1.741	0.081663	.

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 26943 on 19434 degrees of freedom
Residual deviance: 20246 on 19417 degrees of freedom
AIC: 20282

```

```

> anova(moda, modb, modc, modabc, test="Chisq")

```

Analysis of Deviance Table

```

Model 1: esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
esrendtot:esnpagr +
esnpagr:esregocup

```

```

Model 2: esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
esnpagr:esregocup +
esregocup:esgurb

```

```
Model 3: esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
esrendtot:esnpagr +
  esregocup:esgurb
```

```
Model 4: esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
esrendtot:esnpagr +
  esnpagr:esregocup + esnpagr:esregocup + esregocup:esgurb +
esrendtot:esnpagr + esregocup:esgurb
```

	Resid. Df	Resid. Dev	Df	Deviance	P(> Chi)
1	19423	20280			
2	19418	20276	5	4.6284	0.4629
3	19420	20277	-2	-0.8307	0.6601
4	19417	20246	3	30.4208	1.126e-06 ***

```
MODELO FINAL ESPANHA> modabc<-
glm(esdesptot~esrendtot+esnpagr+esregocup+esgurb+esrendtot:esnpagr+esn
pagr:esregocup+esnpagr:esregocup:esgurb+esrendtot:esnpagr+es
regocup:esgurb, family=binomial("logit"))
> summary(modabc)
```

Call:

```
glm(formula = esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
  esrendtot:esnpagr + esnpagr:esregocup + esnpagr:esregocup +
  esregocup:esgurb + esrendtot:esnpagr + esregocup:esgurb,
  family = binomial("logit"))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.9251	-0.7379	0.5839	0.7723	2.3009

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.50674	0.04804	-31.364	< 2e-16 ***
esrendtot1	2.01423	0.05761	34.965	< 2e-16 ***
esnpagr1	1.55077	0.05762	26.912	< 2e-16 ***
esregocup2	0.34498	0.07632	4.520	6.18e-06 ***
esregocup3	-0.76946	0.10868	-7.080	1.44e-12 ***
esregocup4	-0.17319	0.15683	-1.104	0.269465
esgurb2	-0.24910	0.06175	-4.034	5.49e-05 ***
esgurb3	-0.60502	0.05371	-11.264	< 2e-16 ***
esrendtot1:esnpagr1	-0.39600	0.07300	-5.425	5.80e-08 ***
esnpagr1:esregocup2	-0.44349	0.08410	-5.273	1.34e-07 ***
esnpagr1:esregocup3	-0.35453	0.12440	-2.850	0.004373 **
esnpagr1:esregocup4	-0.17168	0.16306	-1.053	0.292395
esregocup2:esgurb2	0.36796	0.10098	3.644	0.000268 ***
esregocup3:esgurb2	0.21429	0.14966	1.432	0.152189
esregocup4:esgurb2	-0.01495	0.20814	-0.072	0.942722
esregocup2:esgurb3	0.48994	0.09348	5.241	1.59e-07 ***
esregocup3:esgurb3	0.30764	0.14887	2.067	0.038780 *
esregocup4:esgurb3	0.31198	0.17918	1.741	0.081663 .

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 26943 on 19434 degrees of freedom
Residual deviance: 20246 on 19417 degrees of freedom
AIC: 20282

Comparação de valores antes e depois de stepwise

```
> final (antes step)
```



```

> modfinal<-
glm(esdesptot~esrendtot+esnpagr+esregocup+esgurb+esrendtot:esnpagr+esn
pagr:esregocup, family=binomial("logit"))
> summary (modfinal)

Call:
glm(formula = esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
     esrendtot:esnpagr + esnpagr:esregocup, family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9371  -0.7382   0.5766   0.7463   2.3376

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.59430    0.04527  -35.216 < 2e-16 ***
esrendtot1         2.02147    0.05754   35.133 < 2e-16 ***
esnpagr1           1.53567    0.05725   26.823 < 2e-16 ***
esregocup2         0.54343    0.06725    8.081 6.45e-16 ***
esregocup3        -0.65173    0.09861   -6.609 3.86e-11 ***
esregocup4        -0.08648    0.12842   -0.673 0.50071
esgurb2           -0.11010    0.04474   -2.461 0.01386 *
esgurb3           -0.41893    0.04055  -10.331 < 2e-16 ***
esrendtot1:esnpagr1 -0.40786    0.07288   -5.596 2.19e-08 ***
esnpagr1:esregocup2 -0.38843    0.08366   -4.643 3.43e-06 ***
esnpagr1:esregocup3 -0.31952    0.12348   -2.588 0.00966 **
esnpagr1:esregocup4 -0.13211    0.16243   -0.813 0.41604
---
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 26943  on 19434  degrees of freedom
Residual deviance: 20280  on 19423  degrees of freedom
AIC: 20304

Number of Fisher Scoring iterations: 4

>
> step
function (object, scope, scale = 0, direction = c("both", "backward",
  "forward"), trace = 1, keep = NULL, steps = 1000, k = 2,
  ...)
{
  mydeviance <- function(x, ...) {
    dev <- deviance(x)
    if (!is.null(dev))
      dev
    else extractAIC(x, k = 0)[2L]
  }
  cut.string <- function(string) {
    if (length(string) > 1)
      string[-1L] <- paste("\n", string[-1L], sep = "")
    string
  }
  re.arrange <- function(keep) {
    namr <- names(k1 <- keep[[1L]])
    namc <- names(keep)
    nc <- length(keep)
    nr <- length(k1)
    array(unlist(keep, recursive = FALSE), c(nr, nc), list(namr,
      namc))
  }
}

```

```

}
step.results <- function(models, fit, object, usingCp = FALSE) {
  change <- sapply(models, "[", "change")
  rd <- sapply(models, "[", "deviance")
  dd <- c(NA, abs(diff(rd)))
  rdf <- sapply(models, "[", "df.resid")
  ddf <- c(NA, diff(rdf))
  AIC <- sapply(models, "[", "AIC")
  heading <- c("Stepwise Model Path \nAnalysis of Deviance
Table",
              "\nInitial Model:", deparse(as.vector(formula(object))),
              "\nFinal Model:", deparse(as.vector(formula(fit))),
              "\n")
  aod <- data.frame(Step = I(change), Df = ddf, Deviance = dd,
                  `Resid. Df` = rdf, `Resid. Dev` = rd, AIC = AIC,
                  check.names = FALSE)
  if (usingCp) {
    cn <- colnames(aod)
    cn[cn == "AIC"] <- "Cp"
    colnames(aod) <- cn
  }
  attr(aod, "heading") <- heading
  fit$anova <- aod
  fit
}
Terms <- terms(object)
object$call$formula <- object$formula <- Terms
md <- missing(direction)
direction <- match.arg(direction)
backward <- direction == "both" | direction == "backward"
forward <- direction == "both" | direction == "forward"
if (missing(scope)) {
  fdrop <- numeric(0L)
  fadd <- attr(Terms, "factors")
  if (md)
    forward <- FALSE
}
else {
  if (is.list(scope)) {
    fdrop <- if (!is.null(fdrop <- scope$lower))
      attr(terms(update.formula(object, fdrop)), "factors")
    else numeric(0L)
    fadd <- if (!is.null(fadd <- scope$upper))
      attr(terms(update.formula(object, fadd)), "factors")
  }
  else {
    fadd <- if (!is.null(fadd <- scope))
      attr(terms(update.formula(object, scope)), "factors")
    fdrop <- numeric(0L)
  }
}
models <- vector("list", steps)
if (!is.null(keep))
  keep.list <- vector("list", steps)
n <- length(object$residuals)
fit <- object
bAIC <- extractAIC(fit, scale, k = k, ...)
edf <- bAIC[1L]
bAIC <- bAIC[2L]
if (is.na(bAIC))

```

```

        stop("AIC is not defined for this model, so 'step' cannot
proceed")
    nm <- 1
    Terms <- fit$terms
    if (trace)
        cat("Start:  AIC=", format(round(bAIC, 2)), "\n",
cut.string(deparse(as.vector(formula(fit)))),
"\n\n", sep = "")
    models[[nm]] <- list(deviance = mydeviance(fit), df.resid = n -
edf, change = "", AIC = bAIC)
    if (!is.null(keep))
        keep.list[[nm]] <- keep(fit, bAIC)
    usingCp <- FALSE
    while (steps > 0) {
        steps <- steps - 1
        AIC <- bAIC
        ffac <- attr(Terms, "factors")
        scope <- factor.scope(ffac, list(add = fadd, drop = fdrop))
        aod <- NULL
        change <- NULL
        if (backward && length(scope$drop)) {
            aod <- dropl(fit, scope$drop, scale = scale, trace =
trace,
                k = k, ...)
            rn <- row.names(aod)
            row.names(aod) <- c(rn[1L], paste("-", rn[-1L], sep = "
"))
            if (any(aod$Df == 0, na.rm = TRUE)) {
                zdf <- aod$Df == 0 & !is.na(aod$Df)
                change <- rev(rownames(aod)[zdf])[1L]
            }
        }
        if (is.null(change)) {
            if (forward && length(scope$add)) {
                aodf <- addl(fit, scope$add, scale = scale, trace =
trace,
                    k = k, ...)
                rn <- row.names(aodf)
                row.names(aodf) <- c(rn[1L], paste("+", rn[-1L],
                    sep = " "))
                aod <- if (is.null(aod))
                    aodf
                else rbind(aod, aodf[-1, , drop = FALSE])
            }
            attr(aod, "heading") <- NULL
            nzdf <- if (!is.null(aod$Df))
                aod$Df != 0 | is.na(aod$Df)
            aod <- aod[nzdf, ]
            if (is.null(aod) || ncol(aod) == 0)
                break
            nc <- match(c("Cp", "AIC"), names(aod))
            nc <- nc[!is.na(nc)][1L]
            o <- order(aod[, nc])
            if (trace)
                print(aod[o, ])
            if (o[1L] == 1)
                break
            change <- rownames(aod)[o[1L]]
        }
    }
    usingCp <- match("Cp", names(aod), 0L) > 0L
    fit <- update(fit, paste("~ .", change), evaluate = FALSE)

```

```

fit <- eval.parent(fit)
if (length(fit$residuals) != n)
  stop("number of rows in use has changed: remove missing
values?")
Terms <- terms(fit)
bAIC <- extractAIC(fit, scale, k = k, ...)
edf <- bAIC[1L]
bAIC <- bAIC[2L]
if (trace)
  cat("\nStep: AIC=", format(round(bAIC, 2)), "\n",
      cut.string(deparse(as.vector(formula(fit))),
                "\n\n", sep = ""))
if (bAIC >= AIC + 1e-07)
  break
nm <- nm + 1
models[[nm]] <- list(deviance = mydeviance(fit), df.resid = n
-
      edf, change = change, AIC = bAIC)
if (!is.null(keep))
  keep.list[[nm]] <- keep(fit, bAIC)
}
if (!is.null(keep))
  fit$keep <- re.arrange(keep.list[seq(nm)])
step.results(models = models[seq(nm)], fit, object, usingCp)
}
<environment: namespace:stats>
> modfinalstep<-
glm(esdesptot~esrendtot+esnpagr+esregocup+esgurb+esrendtot:esnpagr+esr
endtot:esregocup+esrendtot:esgurb+esnpagr:esregocup+esnpagr:esgurb+esr
egocup:esgurb, family=binomial("logit"))
> mod<-step(modfinalstep)
Start: AIC=20292.48
esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
esrendtot:esnpagr +
  esrendtot:esregocup + esrendtot:esgurb + esnpagr:esregocup +
  esnpagr:esgurb + esregocup:esgurb

              Df Deviance  AIC
- esnpagr:esgurb      2    20243 20289
- esrendtot:esgurb    2    20243 20289
- esrendtot:esregocup 3    20245 20289
<none>                20243 20293
- esregocup:esgurb    6    20277 20315
- esnpagr:esregocup   3    20274 20318
- esrendtot:esnpagr   1    20271 20319

Step: AIC=20288.77
esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
esrendtot:esnpagr +
  esrendtot:esregocup + esrendtot:esgurb + esnpagr:esregocup +
  esregocup:esgurb

              Df Deviance  AIC
- esrendtot:esgurb    2    20243 20285
- esrendtot:esregocup 3    20246 20286
<none>                20243 20289
- esregocup:esgurb    6    20277 20311
- esnpagr:esregocup   3    20274 20314
- esrendtot:esnpagr   1    20271 20315

Step: AIC=20285.16

```

```

esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
esrendtot:esnpagr +
  esrendtot:esregocup + esnpagr:esregocup + esregocup:esgurb

              Df Deviance   AIC
- esrendtot:esregocup  3    20246 20282
<none>                20243 20285
- esregocup:esgurb     6    20277 20307
- esnpagr:esregocup    3    20274 20310
- esrendtot:esnpagr    1    20273 20313

Step:  AIC=20282.04
esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
esrendtot:esnpagr +
  esnpagr:esregocup + esregocup:esgurb

              Df Deviance   AIC
<none>                20246 20282
- esregocup:esgurb     6    20280 20304
- esnpagr:esregocup    3    20277 20307
- esrendtot:esnpagr    1    20276 20310

```

```

> summary (modfinalstep)

Call:
glm(formula = esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
esrendtot:esnpagr + esrendtot:esregocup + esrendtot:esgurb +
  esnpagr:esregocup + esnpagr:esgurb + esregocup:esgurb, family =
binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9257  -0.7242   0.5836   0.7830   2.3239

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.508153   0.056082 -26.892 < 2e-16 ***
esrendtot1    2.010236   0.073644  27.297 < 2e-16 ***
esnpagr1      1.556231   0.072888  21.351 < 2e-16 ***
esregocup2    0.352599   0.088965   3.963 7.39e-05 ***
esregocup3   -0.868845   0.127696  -6.804 1.02e-11 ***
esregocup4   -0.184563   0.166511  -1.108 0.267684
esgurb2      -0.262378   0.089122  -2.944 0.003240 **
esgurb3     -0.570294   0.079156  -7.205 5.82e-13 ***
esrendtot1:esnpagr1 -0.393776   0.074228  -5.305 1.13e-07 ***
esrendtot1:esregocup2 -0.003951   0.081287  -0.049 0.961232
esrendtot1:esregocup3  0.192172   0.121867   1.577 0.114819
esrendtot1:esregocup4  0.034891   0.161615   0.216 0.829076
esrendtot1:esgurb2  -0.013363   0.091282  -0.146 0.883615
esrendtot1:esgurb3  -0.045793   0.082977  -0.552 0.581031
esnpagr1:esregocup2 -0.452099   0.084514  -5.349 8.83e-08 ***
esnpagr1:esregocup3 -0.366992   0.126946  -2.891 0.003841 **
esnpagr1:esregocup4 -0.177222   0.164351  -1.078 0.280895
esnpagr1:esgurb2    0.031562   0.093181   0.339 0.734820
esnpagr1:esgurb3   -0.023261   0.085755  -0.271 0.786199
esregocup2:esgurb2  0.366865   0.102107   3.593 0.000327 ***
esregocup3:esgurb2  0.214189   0.152316   1.406 0.159660
esregocup4:esgurb2 -0.016714   0.209086  -0.080 0.936286
esregocup2:esgurb3  0.495723   0.094269   5.259 1.45e-07 ***
esregocup3:esgurb3  0.316541   0.150961   2.097 0.036008 *
esregocup4:esgurb3  0.311003   0.179578   1.732 0.083299 .

```

```

---
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 26943 on 19434 degrees of freedom
Residual deviance: 20242 on 19410 degrees of freedom
AIC: 20292

```

```

> anova(modfinal, modfinalstep, test="Chisq")
Analysis of Deviance Table

Model 1: esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
esrendtot:esnpagr +
  esnpagr:esregocup
Model 2: esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
esrendtot:esnpagr +
  esrendtot:esregocup + esrendtot:esgurb + esnpagr:esregocup +
  esnpagr:esgurb + esregocup:esgurb
  Resid. Df Resid. Dev Df Deviance P(>|Chi|)
1      19423      20280
2      19410      20243 13    37.77 0.0003135 ***

```

```

> mod<- glm(rmonet + rnaomonet + npagr + regocup + conftotal +
+   npagr:conftotal + regocup:conftotal +
+   npagr:gurb + rmonet:npagr + regocup:gurb + rmonet:regocup +
+   rnaomonet:regocup, family=binomial("logit"))
> summary (mod)

Call:
glm(formula = esdesptot ~ esrendtot + esnpagr + esregocup + esgurb +
  esrendtot:esnpagr + esnpagr:esregocup + esregocup:esgurb,
  family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9251  -0.7379   0.5839   0.7723   2.3009

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.50674    0.04804  -31.364 < 2e-16 ***
esrendtot1     2.01423    0.05761   34.965 < 2e-16 ***
esnpagr1       1.55077    0.05762   26.912 < 2e-16 ***
esregocup2     0.34498    0.07632    4.520 6.18e-06 ***
esregocup3    -0.76946    0.10868   -7.080 1.44e-12 ***
esregocup4    -0.17319    0.15683   -1.104 0.269465
esgurb2       -0.24910    0.06175   -4.034 5.49e-05 ***
esgurb3       -0.60502    0.05371  -11.264 < 2e-16 ***
esrendtot1:esnpagr1 -0.39600    0.07300   -5.425 5.80e-08 ***
esnpagr1:esregocup2 -0.44349    0.08410   -5.273 1.34e-07 ***
esnpagr1:esregocup3 -0.35453    0.12440   -2.850 0.004373 **
esnpagr1:esregocup4 -0.17168    0.16306   -1.053 0.292395
esregocup2:esgurb2  0.36796    0.10098    3.644 0.000268 ***
esregocup3:esgurb2  0.21429    0.14966    1.432 0.152189
esregocup4:esgurb2 -0.01495    0.20814   -0.072 0.942722
esregocup2:esgurb3  0.48994    0.09348    5.241 1.59e-07 ***
esregocup3:esgurb3  0.30764    0.14887    2.067 0.038780 *
esregocup4:esgurb3  0.31198    0.17918    1.741 0.081663 .
---
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 26943 on 19434 degrees of freedom

```

```
Residual deviance: 20246 on 19417 degrees of freedom
AIC: 20282
```

	Df	Deviance	Resid. Df	Resid. Dev	P(> Chi)
NULL			19434	26943	
esrendtot	1	5020.6	19433	21922	< 2.2e-16 ***
esnpagr	1	1132.7	19432	20789	< 2.2e-16 ***
esregocup	3	330.4	19429	20459	< 2.2e-16 ***
esgurb	2	113.0	19427	20346	< 2.2e-16 ***
esrendtot:esnpagr	1	41.8	19426	20304	1.003e-10 ***

```
> modfinal<-
glm(esdesptot~esrendtot+esnpagr+esregocup_novo+esgurb+esrendtot:esregocup_novo+esrendtot:esgurb+esnpagr:esregocup_novo+esregocup_novo:esgurb, family=binomial("logit"))
> summary(modfinal)
```

```
Call:
glm(formula = esdesptot ~ esrendtot + esnpagr + esregocup + esgurb + esrendtot:esnpagr + esnpagr:esregocup, family = binomial("logit"))
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9371  -0.7382   0.5766   0.7463   2.3376
```

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.59430    0.04527  -35.216 < 2e-16 ***
esrendtot1     2.02147    0.05754   35.133 < 2e-16 ***
esnpagr1       1.53567    0.05725   26.823 < 2e-16 ***
esregocup2     0.54343    0.06725    8.081 6.45e-16 ***
esregocup3    -0.65173    0.09861   -6.609 3.86e-11 ***
esregocup4    -0.08648    0.12842   -0.673  0.50071
esgurb2       -0.11010    0.04474   -2.461  0.01386 *
esgurb3       -0.41893    0.04055  -10.331 < 2e-16 ***
esrendtot1:esnpagr1 -0.40786    0.07288   -5.596 2.19e-08 ***
esnpagr1:esregocup2 -0.38843    0.08366   -4.643 3.43e-06 ***
esnpagr1:esregocup3 -0.31952    0.12348   -2.588 0.00966 **
esnpagr1:esregocup4 -0.13211    0.16243   -0.813  0.41604
```

```
---
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 26943 on 19434 degrees of freedom
Residual deviance: 20280 on 19423 degrees of freedom
AIC: 20304
```

Matriz de covariância

```
> vcov(modfinal)
              (Intercept)    esrendtot1    esnpagr1
esregocup2
(Intercept)    0.0020495052 -1.218704e-03 -0.0016078819 -
1.021896e-03
esrendtot1    -0.0012187042  3.310512e-03  0.0011482375 -
8.205671e-04
esnpagr1      -0.0016078819  1.148237e-03  0.0032777861
9.672863e-04
esregocup2    -0.0010218958 -8.205671e-04  0.0009672863
4.522681e-03
```

esregocup3	-0.0011698115	-5.659592e-04	0.0010411691	
1.432701e-03				
esregocup4	-0.0012642909	6.010500e-05	0.0012944681	
1.252507e-03				
esgurb2	-0.0006546781	6.813607e-05	-0.0000918810	
2.833154e-05				
esgurb3	-0.0006963943	1.311626e-04	-0.0001476988	
1.151498e-04				
esrendtot1:esnpagr1	0.0011802627	-3.303966e-03	-0.0022889362	
8.259177e-04				
esnpagr1:esregocup2	0.0009831936	8.294132e-04	-0.0018870154	-
4.514026e-03				
esnpagr1:esregocup3	0.0010739074	5.848406e-04	-0.0020199889	-
1.415658e-03				
esnpagr1:esregocup4	0.0012851346	-6.458435e-05	-0.0023826483	-
1.256756e-03				
	esregocup3	esregocup4	esgurb2	
esgurb3				
(Intercept)	-0.0011698115	-1.264291e-03	-6.546781e-04	-
6.963943e-04				
esrendtot1	-0.0005659592	6.010500e-05	6.813607e-05	
1.311626e-04				
esnpagr1	0.0010411691	1.294468e-03	-9.188099e-05	-
1.476988e-04				
esregocup2	0.0014327008	1.252507e-03	2.833154e-05	
1.151498e-04				
esregocup3	0.0097234257	1.253429e-03	1.224997e-04	
2.404893e-04				
esregocup4	0.0012534287	1.649248e-02	-8.885614e-05	-
2.324185e-05				
esgurb2	0.0001224997	-8.885614e-05	2.001722e-03	
6.834109e-04				
esgurb3	0.0002404893	-2.324185e-05	6.834109e-04	
1.644402e-03				
esrendtot1:esnpagr1	0.0005779308	-6.223122e-05	-1.330570e-05	-
5.124452e-05				
esnpagr1:esregocup2	-0.0014164126	-1.251908e-03	-2.870654e-05	-
5.758753e-07				
esnpagr1:esregocup3	-0.0096887704	-1.255636e-03	-4.854312e-05	-
1.238131e-06				
esnpagr1:esregocup4	-0.0012616660	-1.649246e-02	8.203873e-05	-
3.463012e-05				
	esrendtot1:esnpagr1	esnpagr1:esregocup2		
esnpagr1:esregocup3				
(Intercept)	1.180263e-03	9.831936e-04		
1.073907e-03				
esrendtot1	-3.303966e-03	8.294132e-04		
5.848406e-04				
esnpagr1	-2.288936e-03	-1.887015e-03		-
2.019989e-03				
esregocup2	8.259177e-04	-4.514026e-03		-
1.415658e-03				
esregocup3	5.779308e-04	-1.416413e-03		-
9.688770e-03				
esregocup4	-6.223122e-05	-1.251908e-03		-
1.255636e-03				
esgurb2	-1.330570e-05	-2.870654e-05		-
4.854312e-05				
esgurb3	-5.124452e-05	5.758753e-07		-
1.238131e-06				

esrendtot1:esnpagr1	5.312162e-03	-8.976139e-04	-
5.689322e-04			
esnpagr1:esregocup2	-8.976139e-04	6.999040e-03	
2.364809e-03			
esnpagr1:esregocup3	-5.689322e-04	2.364809e-03	
1.524638e-02			
esnpagr1:esregocup4	3.207391e-04	2.191378e-03	
2.205755e-03			
	esnpagr1:esregocup4		
(Intercept)	1.285135e-03		
esrendtot1	-6.458435e-05		
esnpagr1	-2.382648e-03		
esregocup2	-1.256756e-03		
esregocup3	-1.261666e-03		
esregocup4	-1.649246e-02		
esgurb2	8.203873e-05		
esgurb3	-3.463012e-05		
esrendtot1:esnpagr1	3.207391e-04		
esnpagr1:esregocup2	2.191378e-03		
esnpagr1:esregocup3	2.205755e-03		
esnpagr1:esregocup4	2.638443e-02		

```

> Ponto de corte

> tp<-sum(esdesptot==1)
> tp
[1] 9719
> vp<-sum((modfinal$fitted>=0.50) & (esdesptot==1))
> vp
[1] 7355
> sensb<-vp/tp
> sensb*100
[1] 75.67651
> tn<-sum(esdesptot==0)
> tn
[1] 9716
> vn<-sum((modfinal$fitted<0.50) & (esdesptot==0))
> vn
[1] 7223
> esp<-vn/tn
> esp*100
[1] 74.3413

```

Anexo IV

Modelo de Portugal e Espanha (scripts)

```
glm(formula = desptot ~ rendtot + npagr + regocup + gurb + pais,
     family = binomial("logit"))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.0954	-0.6721	0.4858	0.7448	2.3124

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.37279	0.03443	-39.871	< 2e-16	***
rendtot1	1.94707	0.02905	67.020	< 2e-16	***
npagr1	1.14489	0.02937	38.984	< 2e-16	***
regocup2	0.35810	0.03468	10.325	< 2e-16	***
regocup3	-0.65092	0.04569	-14.245	< 2e-16	***
regocup4	-0.17673	0.06281	-2.814	0.0049	**
gurb2	-0.20540	0.03700	-5.551	2.84e-08	***
gurb3	-0.49909	0.03501	-14.257	< 2e-16	***
pais1	-0.07942	0.03085	-2.574	0.0100	*

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 41364 on 29837 degrees of freedom
Residual deviance: 30297 on 29829 degrees of freedom
AIC: 30315

```
> modi<-glm(desptot~rendtot+npagr+regocup+gurb+pais+pais:rendtot,
             family=binomial("logit"))
> summary (modi)
```

Call:

```
glm(formula = desptot ~ rendtot + npagr + regocup + gurb + pais +
     pais:rendtot, family = binomial("logit"))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.1589	-0.6702	0.4524	0.7276	2.3366

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.53731	0.04133	-37.196	< 2e-16	***
rendtot1	2.26788	0.05124	44.256	< 2e-16	***
npagr1	1.14218	0.02940	38.845	< 2e-16	***
regocup2	0.35527	0.03462	10.263	< 2e-16	***
regocup3	-0.63349	0.04597	-13.780	< 2e-16	***
regocup4	-0.17199	0.06301	-2.729	0.006344	**
gurb2	-0.19518	0.03705	-5.268	1.38e-07	***
gurb3	-0.49157	0.03492	-14.076	< 2e-16	***
pais1	0.15810	0.04396	3.597	0.000322	***
rendtot1:pais1	-0.47483	0.06141	-7.732	1.06e-14	***

Null deviance: 41364 on 29837 degrees of freedom
Residual deviance: 30236 on 29828 degrees of freedom
AIC: 30256

```

> modii<-glm(desptot~rendtot+npagr+regocup+gurb+pais+pais:npagr,
family=binomial("logit"))
> summary (modii)

Call:
glm(formula = desptot ~ rendtot + npagr + regocup + gurb + pais +
  pais:npagr, family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0841  -0.6803   0.4920   0.7414   2.3201

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.34561    0.03931  -34.228 < 2e-16 ***
rendtot1      1.94769    0.02906   67.022 < 2e-16 ***
npagr1        1.08898    0.04912   22.171 < 2e-16 ***
regocup2      0.35969    0.03471   10.363 < 2e-16 ***
regocup3     -0.64995    0.04568  -14.228 < 2e-16 ***
regocup4     -0.17505    0.06281   -2.787  0.00532 **
gurb2        -0.20556    0.03700   -5.556 2.77e-08 ***
gurb3        -0.50145    0.03506  -14.301 < 2e-16 ***
pais1        -0.12423    0.04414   -2.814  0.00489 **
npagr1:pais1  0.08565    0.06041    1.418  0.15627

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 41364 on 29837 degrees of freedom
Residual deviance: 30295 on 29828 degrees of freedom
AIC: 30315

```

```

> modiii<-glm(desptot~rendtot+npagr+regocup+gurb+pais+pais:gurb,
family=binomial("logit"))
> summary (modiii)

Call:
glm(formula = desptot ~ rendtot + npagr + regocup + gurb + pais +
  pais:gurb, family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1276  -0.6985   0.4687   0.7296   2.3559

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.28638    0.03808  -33.780 < 2e-16 ***
rendtot1      1.94418    0.02907   66.887 < 2e-16 ***
npagr1        1.14249    0.02940   38.858 < 2e-16 ***
regocup2      0.35323    0.03470   10.180 < 2e-16 ***
regocup3     -0.66944    0.04583  -14.606 < 2e-16 ***
regocup4     -0.18342    0.06287   -2.917  0.003530 **
gurb2        -0.39920    0.06430   -6.208 5.35e-10 ***
gurb3        -0.75493    0.06825  -11.061 < 2e-16 ***
pais1        -0.21567    0.04035   -5.345 9.02e-08 ***
gurb2:pais1  0.29906    0.07834    3.818 0.000135 ***
gurb3:pais1  0.35800    0.07900    4.532 5.85e-06 ***
---
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 41364 on 29837 degrees of freedom
Residual deviance: 30269 on 29827 degrees of freedom

```

```
AIC: 30291
```

```
> modiv<-glm(desptot~rendtot+npagr+regocup+gurb+pais+pais:regocup,  
family=binomial("logit"))  
> summary (modiv)
```

Call:

```
glm(formula = desptot ~ rendtot + npagr + regocup + gurb + pais +  
    pais:regocup, family = binomial("logit"))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.1203	-0.6564	0.4725	0.7337	2.3783

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.43673	0.03843	-37.382	< 2e-16	***
rendtot1	1.95439	0.02917	67.005	< 2e-16	***
npagr1	1.14104	0.02941	38.799	< 2e-16	***
regocup2	0.47740	0.06583	7.252	4.09e-13	***
regocup3	-0.39137	0.06869	-5.698	1.21e-08	***
regocup4	-0.24706	0.10217	-2.418	0.0156	*
gurb2	-0.19437	0.03713	-5.236	1.65e-07	***
gurb3	-0.49498	0.03508	-14.111	< 2e-16	***
pais1	0.01120	0.03987	0.281	0.7788	
regocup2:pais1	-0.17191	0.07682	-2.238	0.0252	*
regocup3:pais1	-0.45534	0.09144	-4.980	6.36e-07	***
regocup4:pais1	0.12025	0.12940	0.929	0.3528	

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 41364 on 29837 degrees of freedom
Residual deviance: 30268 on 29826 degrees of freedom
AIC: 30292

```
> modv<-glm(desptot~rendtot+npagr+regocup+gurb+pais+rendtot:npagr,  
family=binomial("logit"))  
> summary (modv)
```

Call:

```
glm(formula = desptot ~ rendtot + npagr + regocup + gurb + pais +  
    rendtot:npagr, family = binomial("logit"))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.0389	-0.6299	0.5170	0.7518	2.3805

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.51686	0.03787	-40.057	< 2e-16	***
rendtot1	2.30970	0.04534	50.941	< 2e-16	***
npagr1	1.45463	0.04164	34.935	< 2e-16	***
regocup2	0.32628	0.03473	9.396	< 2e-16	***
regocup3	-0.68102	0.04575	-14.886	< 2e-16	***
regocup4	-0.19601	0.06295	-3.114	0.00185	**
gurb2	-0.19788	0.03705	-5.342	9.21e-08	***
gurb3	-0.48897	0.03504	-13.955	< 2e-16	***
pais1	-0.08593	0.03098	-2.773	0.00555	**
rendtot1:npagr1	-0.62881	0.05882	-10.691	< 2e-16	***

Null deviance: 41364 on 29837 degrees of freedom
Residual deviance: 30182 on 29828 degrees of freedom

```
AIC: 30202
```

```
> modvi<-glm(desptot~rendtot+npagr+regocup+gurb+pais+rendtot:regocup,  
family=binomial("logit"))  
> summary (modvi)
```

Call:

```
glm(formula = desptot ~ rendtot + npagr + regocup + gurb + pais +  
rendtot:regocup, family = binomial("logit"))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.0708	-0.6644	0.4992	0.7348	2.3089

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.39858	0.03625	-38.581	< 2e-16	***
rendtot1	2.01345	0.03863	52.117	< 2e-16	***
npagr1	1.13544	0.02956	38.416	< 2e-16	***
regocup2	0.47881	0.05286	9.059	< 2e-16	***
regocup3	-0.61572	0.06258	-9.839	< 2e-16	***
regocup4	-0.13859	0.08133	-1.704	0.08835	.
gurb2	-0.20355	0.03703	-5.496	3.88e-08	***
gurb3	-0.49704	0.03507	-14.174	< 2e-16	***
pais1	-0.08199	0.03093	-2.651	0.00802	**
rendtot1:regocup2	-0.20964	0.06934	-3.024	0.00250	**
rendtot1:regocup3	-0.07295	0.09076	-0.804	0.42156	
rendtot1:regocup4	-0.08192	0.12774	-0.641	0.52133	

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 41364 on 29837 degrees of freedom
Residual deviance: 30288 on 29826 degrees of freedom
AIC: 30312

```
> modvii<-glm(desptot~rendtot+npagr+regocup+gurb+pais+rendtot:gurb,  
family=binomial("logit"))  
> summary (modvii)
```

Call:

```
glm(formula = desptot ~ rendtot + npagr + regocup + gurb + pais +  
rendtot:gurb, family = binomial("logit"))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.1067	-0.6616	0.4798	0.7579	2.2937

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.40798	0.03820	-36.862	< 2e-16	***
rendtot1	2.00683	0.04022	49.899	< 2e-16	***
npagr1	1.14756	0.02940	39.028	< 2e-16	***
regocup2	0.35754	0.03470	10.305	< 2e-16	***
regocup3	-0.64601	0.04585	-14.089	< 2e-16	***
regocup4	-0.17480	0.06278	-2.784	0.00536	**
gurb2	-0.16126	0.05198	-3.102	0.00192	**
gurb3	-0.42475	0.04767	-8.910	< 2e-16	***
pais1	-0.07696	0.03091	-2.490	0.01278	*
rendtot1:gurb2	-0.08593	0.07332	-1.172	0.24124	
rendtot1:gurb3	-0.15391	0.06778	-2.271	0.02316	*

```

---
Null deviance: 41364 on 29837 degrees of freedom
Residual deviance: 30292 on 29827 degrees of freedom
AIC: 30314

```

```

> modviii<-glm(desptot~rendtot+npagr+regocup+gurb+pais+npagr:gurb,
family=binomial("logit"))
> summary (modviii)

```

Call:

```

glm(formula = desptot ~ rendtot + npagr + regocup + gurb + pais +
  npagr:gurb, family = binomial("logit"))

```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.0799	-0.6845	0.4942	0.7320	2.3395

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.33188	0.03733	-35.682	< 2e-16	***
rendtot1	1.94420	0.02907	66.885	< 2e-16	***
npagr1	1.07129	0.03962	27.042	< 2e-16	***
regocup2	0.35731	0.03466	10.308	< 2e-16	***
regocup3	-0.65266	0.04561	-14.308	< 2e-16	***
regocup4	-0.17858	0.06285	-2.841	0.00449	**
gurb2	-0.27772	0.05658	-4.909	9.16e-07	***
gurb3	-0.60324	0.05367	-11.239	< 2e-16	***
pais1	-0.08194	0.03086	-2.655	0.00793	**
npagr1:gurb2	0.13210	0.07434	1.777	0.07557	.
npagr1:gurb3	0.18181	0.06943	2.619	0.00883	**

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 41364 on 29837 degrees of freedom
Residual deviance: 30289 on 29827 degrees of freedom
AIC: 30311

```

```

> modix<-glm(desptot~rendtot+npagr+regocup+gurb+pais+npagr:regocup,
family=binomial("logit"))
> summary (modix)

```

Call:

```

glm(formula = desptot ~ rendtot + npagr + regocup + gurb + pais +
  npagr:regocup, family = binomial("logit"))

```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.0502	-0.6522	0.5107	0.7232	2.2728

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.43974	0.03628	-39.683	< 2e-16	***
rendtot1	1.92840	0.02916	66.130	< 2e-16	***
npagr1	1.30647	0.03874	33.723	< 2e-16	***
regocup2	0.63938	0.05547	11.527	< 2e-16	***
regocup3	-0.47700	0.06823	-6.991	2.74e-12	***
regocup4	-0.10344	0.09783	-1.057	0.290332	
gurb2	-0.20273	0.03706	-5.470	4.49e-08	***
gurb3	-0.49559	0.03510	-14.119	< 2e-16	***
pais1	-0.09193	0.03096	-2.969	0.002989	**
npagr1:regocup2	-0.46314	0.07035	-6.584	4.59e-11	***

```

npagr1:regocup3 -0.32064    0.09050   -3.543 0.000396 ***
npagr1:regocup4 -0.14962    0.12759   -1.173 0.240902

```

```

Null deviance: 41364 on 29837 degrees of freedom
Residual deviance: 30249 on 29826 degrees of freedom
AIC: 30273

```

```

> modx<-glm(desptot~rendtot+npagr+regocup+gurb+pais+gurb:regocup,
family=binomial("logit"))
> summary (modx)

```

Call:

```

glm(formula = desptot ~ rendtot + npagr + regocup + gurb + pais +
gurb:regocup, family = binomial("logit"))

```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.0582	-0.6918	0.5062	0.7712	2.2935

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.30799	0.03704	-35.316	< 2e-16	***
rendtot1	1.94753	0.02910	66.935	< 2e-16	***
npagr1	1.14355	0.02940	38.891	< 2e-16	***
regocup2	0.20691	0.04729	4.375	1.21e-05	***
regocup3	-0.72967	0.05662	-12.888	< 2e-16	***
regocup4	-0.34129	0.09138	-3.735	0.000188	***
gurb2	-0.29085	0.04883	-5.956	2.59e-09	***
gurb3	-0.63723	0.04433	-14.374	< 2e-16	***
pais1	-0.08533	0.03094	-2.758	0.005817	**
regocup2:gurb2	0.24180	0.08664	2.791	0.005259	**
regocup3:gurb2	0.13528	0.12180	1.111	0.266707	
regocup4:gurb2	0.10343	0.16141	0.641	0.521643	
regocup2:gurb3	0.37996	0.08267	4.596	4.31e-06	***
regocup3:gurb3	0.20490	0.12717	1.611	0.107143	
regocup4:gurb3	0.46702	0.14497	3.222	0.001275	**

```

Null deviance: 41364 on 29837 degrees of freedom
Residual deviance: 30267 on 29823 degrees of freedom
AIC: 30297

```

```

> step
function (object, scope, scale = 0, direction = c("both", "backward",
"forward"), trace = 1, keep = NULL, steps = 1000, k = 2,
...)
{
  mydeviance <- function(x, ...) {
    dev <- deviance(x)
    if (!is.null(dev))
      dev
    else extractAIC(x, k = 0)[2L]
  }
  cut.string <- function(string) {
    if (length(string) > 1)
      string[-1L] <- paste("\n", string[-1L], sep = "")
    string
  }
  re.arrange <- function(keep) {
    namr <- names(k1 <- keep[[1L]])
    namc <- names(keep)
    nc <- length(keep)
  }
}

```

```

    nr <- length(k1)
    array(unlist(keep, recursive = FALSE), c(nr, nc), list(namr,
      namc))
  }
  step.results <- function(models, fit, object, usingCp = FALSE) {
    change <- sapply(models, "[", "change")
    rd <- sapply(models, "[", "deviance")
    dd <- c(NA, abs(diff(rd)))
    rdf <- sapply(models, "[", "df.resid")
    ddf <- c(NA, diff(rdf))
    AIC <- sapply(models, "[", "AIC")
    heading <- c("Stepwise Model Path \nAnalysis of Deviance
Table",
      "\nInitial Model:", deparse(as.vector(formula(object))),
      "\nFinal Model:", deparse(as.vector(formula(fit))),
      "\n")
    aod <- data.frame(Step = I(change), Df = ddf, Deviance = dd,
      `Resid. Df` = rdf, `Resid. Dev` = rd, AIC = AIC,
      check.names = FALSE)
    if (usingCp) {
      cn <- colnames(aod)
      cn[cn == "AIC"] <- "Cp"
      colnames(aod) <- cn
    }
    attr(aod, "heading") <- heading
    fit$anova <- aod
    fit
  }
  Terms <- terms(object)
  object$call$formula <- object$formula <- Terms
  md <- missing(direction)
  direction <- match.arg(direction)
  backward <- direction == "both" | direction == "backward"
  forward <- direction == "both" | direction == "forward"
  if (missing(scope)) {
    fdrop <- numeric(0L)
    fadd <- attr(Terms, "factors")
    if (md)
      forward <- FALSE
  }
  else {
    if (is.list(scope)) {
      fdrop <- if (!is.null(fdrop <- scope$lower))
        attr(terms(update.formula(object, fdrop)), "factors")
      else numeric(0L)
      fadd <- if (!is.null(fadd <- scope$upper))
        attr(terms(update.formula(object, fadd)), "factors")
    }
    else {
      fadd <- if (!is.null(fadd <- scope))
        attr(terms(update.formula(object, scope)), "factors")
      fdrop <- numeric(0L)
    }
  }
  models <- vector("list", steps)
  if (!is.null(keep))
    keep.list <- vector("list", steps)
  n <- length(object$residuals)
  fit <- object
  bAIC <- extractAIC(fit, scale, k = k, ...)
  edf <- bAIC[1L]

```



```

bAIC <- bAIC[2L]
if (is.na(bAIC))
  stop("AIC is not defined for this model, so 'step' cannot
proceed")
nm <- 1
Terms <- fit$terms
if (trace)
  cat("Start: AIC=", format(round(bAIC, 2)), "\n",
cut.string(deparse(as.vector(formula(fit))),
"\n\n", sep = ""))
models[[nm]] <- list(deviance = mydeviance(fit), df.resid = n -
edf, change = "", AIC = bAIC)
if (!is.null(keep))
  keep.list[[nm]] <- keep(fit, bAIC)
usingCp <- FALSE
while (steps > 0) {
  steps <- steps - 1
  AIC <- bAIC
  ffac <- attr(Terms, "factors")
  scope <- factor.scope(ffac, list(add = fadd, drop = fdrop))
  aod <- NULL
  change <- NULL
  if (backward && length(scope$drop)) {
    aod <- dropl(fit, scope$drop, scale = scale, trace =
trace,
      k = k, ...)
    rn <- row.names(aod)
    row.names(aod) <- c(rn[1L], paste("-", rn[-1L], sep = "
"))
    if (any(aod$Df == 0, na.rm = TRUE)) {
      zdf <- aod$Df == 0 & !is.na(aod$Df)
      change <- rev(rownames(aod)[zdf])[1L]
    }
  }
  if (is.null(change)) {
    if (forward && length(scope$add)) {
      aodf <- addl(fit, scope$add, scale = scale, trace =
trace,
      k = k, ...)
      rn <- row.names(aodf)
      row.names(aodf) <- c(rn[1L], paste("+", rn[-1L],
      sep = " "))
      aod <- if (is.null(aod))
      aodf
      else rbind(aod, aodf[-1, , drop = FALSE])
    }
    attr(aod, "heading") <- NULL
    nzdf <- if (!is.null(aod$Df))
      aod$Df != 0 | is.na(aod$Df)
    aod <- aod[nzdf, ]
    if (is.null(aod) || ncol(aod) == 0)
      break
    nc <- match(c("Cp", "AIC"), names(aod))
    nc <- nc[!is.na(nc)][1L]
    o <- order(aod[, nc])
    if (trace)
      print(aod[o, ])
    if (o[1L] == 1)
      break
    change <- rownames(aod)[o[1L]]
  }
}

```

```

usingCp <- match("Cp", names(aod), 0L) > 0L
fit <- update(fit, paste("~ .", change), evaluate = FALSE)
fit <- eval.parent(fit)
if (length(fit$residuals) != n)
  stop("number of rows in use has changed: remove missing
values?")
Terms <- terms(fit)
bAIC <- extractAIC(fit, scale, k = k, ...)
edf <- bAIC[1L]
bAIC <- bAIC[2L]
if (trace)
  cat("\nStep:  AIC=", format(round(bAIC, 2)), "\n",
      cut.string(deparse(as.vector(formula(fit))),
                "\n\n", sep = ""))
if (bAIC >= AIC + 1e-07)
  break
nm <- nm + 1
models[[nm]] <- list(deviance = mydeviance(fit), df.resid = n
-
  edf, change = change, AIC = bAIC)
if (!is.null(keep))
  keep.list[[nm]] <- keep(fit, bAIC)
}
if (!is.null(keep))
  fit$keep <- re.arrange(keep.list[seq(nm)])
step.results(models = models[seq(nm)], fit, object, usingCp)
}
<environment: namespace:stats>
> modfinalstep<-
glm(desptot~rendtot+npagr+regocup+gurb+pais+pais:rendtot+pais:npagr+pa
is:gurb+pais:regocup+rendtot:npagr+rendtot:regocup+rendtot:gurb+npagr:
regocup+npagr:gurb+gurb:regocup, family=binomial("logit"))
> mod<-step (modfinalstep)
Start:  AIC=30077.76
desptot ~ rendtot + npagr + regocup + gurb + pais + pais:rendtot +
  pais:npagr + pais:gurb + pais:regocup + rendtot:npagr +
rendtot:regocup +
  rendtot:gurb + npagr:regocup + npagr:gurb + gurb:regocup

              Df Deviance  AIC
- rendtot:regocup  3    30012 30072
- npagr:gurb      2    30012 30074
- rendtot:gurb    2    30012 30074
<none>           30012 30078
- gurb:pais       2    30020 30082
- npagr:pais      1    30019 30083
- regocup:pais    3    30041 30101
- regocup:gurb    6    30049 30103
- npagr:regocup   3    30052 30112
- rendtot:pais    1    30056 30120
- rendtot:npagr   1    30089 30153

Step:  AIC=30072.17
desptot ~ rendtot + npagr + regocup + gurb + pais + rendtot:pais +
  npagr:pais + gurb:pais + regocup:pais + rendtot:npagr +
rendtot:gurb +
  npagr:regocup + npagr:gurb + regocup:gurb

              Df Deviance  AIC
- npagr:gurb      2    30012 30068
- rendtot:gurb    2    30013 30069

```

```

<none>                30012 30072
- gurb:pais           2    30020 30076
- npagr:pais          1    30020 30078
- regocup:pais        3    30042 30096
- regocup:gurb        6    30050 30098
- npagr:regocup       3    30053 30107
- rendtot:pais        1    30058 30116
- rendtot:npagr       1    30091 30149

Step:  AIC=30068.26
desptot ~ rendtot + npagr + regocup + gurb + pais + rendtot:pais +
  npagr:pais + gurb:pais + regocup:pais + rendtot:npagr +
  rendtot:gurb +
  npagr:regocup + regocup:gurb

              Df Deviance   AIC
- rendtot:gurb  2    30013 30065
<none>                30012 30068
- gurb:pais           2    30020 30072
- npagr:pais          1    30020 30074
- regocup:pais        3    30042 30092
- regocup:gurb        6    30051 30095
- npagr:regocup       3    30054 30104
- rendtot:pais        1    30058 30112
- rendtot:npagr       1    30092 30146

Step:  AIC=30064.8
desptot ~ rendtot + npagr + regocup + gurb + pais + rendtot:pais +
  npagr:pais + gurb:pais + regocup:pais + rendtot:npagr +
  npagr:regocup +
  regocup:gurb

              Df Deviance   AIC
<none>                30013 30065
- gurb:pais           2    30021 30069
- npagr:pais          1    30021 30071
- regocup:pais        3    30042 30088
- regocup:gurb        6    30051 30091
- npagr:regocup       3    30054 30100
- rendtot:pais        1    30062 30112
- rendtot:npagr       1    30095 30145
> summary (modfinalstep)

```

```

Call:
glm(formula = desptot ~ rendtot + npagr + regocup + gurb + pais +
  pais:rendtot + pais:npagr + pais:gurb + pais:regocup +
  rendtot:npagr +
  rendtot:regocup + rendtot:gurb + npagr:regocup + npagr:gurb +
  gurb:regocup, family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0474  -0.6176   0.5122   0.7275   2.3020

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.61651    0.05927  -27.274 < 2e-16 ***
rendtot1      2.55726    0.06833   37.426 < 2e-16 ***
npagr1        1.43521    0.07065   20.313 < 2e-16 ***
regocup2      0.60247    0.09495    6.345 2.22e-10 ***

```

regocup3	-0.33609	0.09536	-3.524	0.000425	***
regocup4	-0.21886	0.14619	-1.497	0.134355	
gurb2	-0.40471	0.08974	-4.510	6.48e-06	***
gurb3	-0.76409	0.08806	-8.677	< 2e-16	***
pais1	0.05629	0.06665	0.845	0.398371	
rendtot1:pais1	-0.42936	0.06523	-6.582	4.63e-11	***
npagr1:pais1	0.18089	0.06571	2.753	0.005906	**
gurb2:pais1	0.18096	0.08177	2.213	0.026891	*
gurb3:pais1	0.19510	0.08318	2.345	0.019007	*
regocup2:pais1	-0.20258	0.08075	-2.509	0.012111	*
regocup3:pais1	-0.51714	0.09918	-5.214	1.85e-07	***
regocup4:pais1	-0.01453	0.13559	-0.107	0.914680	
rendtot1:npagr1	-0.53413	0.06085	-8.779	< 2e-16	***
rendtot1:regocup2	-0.03563	0.07067	-0.504	0.614110	
rendtot1:regocup3	0.02110	0.09324	0.226	0.820958	
rendtot1:regocup4	0.01265	0.12888	0.098	0.921831	
rendtot1:gurb2	-0.02529	0.07602	-0.333	0.739327	
rendtot1:gurb3	-0.04933	0.07185	-0.687	0.492332	
npagr1:regocup2	-0.44387	0.07313	-6.070	1.28e-09	***
npagr1:regocup3	-0.29983	0.09480	-3.163	0.001563	**
npagr1:regocup4	-0.23467	0.13100	-1.791	0.073244	.
npagr1:gurb2	0.01516	0.07777	0.195	0.845428	
npagr1:gurb3	-0.01281	0.07429	-0.172	0.863109	
regocup2:gurb2	0.26341	0.08858	2.974	0.002944	**
regocup3:gurb2	0.25269	0.12427	2.033	0.042022	*
regocup4:gurb2	0.12166	0.16360	0.744	0.457080	
regocup2:gurb3	0.42744	0.08487	5.037	4.74e-07	***
regocup3:gurb3	0.40625	0.13034	3.117	0.001828	**
regocup4:gurb3	0.46353	0.15047	3.080	0.002067	**

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 41364 on 29837 degrees of freedom
Residual deviance: 30012 on 29805 degrees of freedom
AIC: 30078

```

MODELO FINAL CONJUNTO > modf<-
glm(desptot~rendtot+npagr+regocup+gurb+pais+pais:rendtot+pais:npagr+pa
is:gurb+pais:regocup+rendtot:npagr+npagr:regocup+gurb:regocup,
family=binomial("logit"))
>
> summary (modf)
Call:
glm(formula = desptot ~ rendtot + npagr + regocup + gurb + pais +
  pais:rendtot + pais:npagr + pais:gurb + pais:regocup +
  rendtot:npagr +
  npagr:regocup + gurb:regocup, family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0495  -0.6213   0.5111   0.7354   2.3061

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.60980    0.05599  -28.752 < 2e-16 ***
rendtot1      2.54578    0.06125  41.564 < 2e-16 ***
npagr1        1.43753    0.06505  22.100 < 2e-16 ***
regocup2      0.57930    0.08467   6.842 7.83e-12 ***
regocup3     -0.33105    0.08821  -3.753 0.000175 ***
regocup4     -0.21513    0.13766  -1.563 0.118105
gurb2        -0.40812    0.07313  -5.581 2.40e-08 ***

```

gurb3	-0.79371	0.07560	-10.499	< 2e-16	***
pais1	0.06282	0.06605	0.951	0.341511	
rendtot1:pais1	-0.44342	0.06355	-6.977	3.01e-12	***
npagr1:pais1	0.18129	0.06466	2.804	0.005048	**
gurb2:pais1	0.18047	0.08172	2.208	0.027220	*
gurb3:pais1	0.19271	0.08320	2.316	0.020542	*
regocup2:pais1	-0.19878	0.08075	-2.461	0.013836	*
regocup3:pais1	-0.51100	0.09795	-5.217	1.82e-07	***
regocup4:pais1	-0.01509	0.13531	-0.111	0.911222	
rendtot1:npagr1	-0.54034	0.05986	-9.027	< 2e-16	***
npagr1:regocup2	-0.44296	0.07264	-6.098	1.07e-09	***
npagr1:regocup3	-0.29792	0.09343	-3.189	0.001430	**
npagr1:regocup4	-0.23405	0.12954	-1.807	0.070801	.
regocup2:gurb2	0.26346	0.08745	3.013	0.002591	**
regocup3:gurb2	0.25417	0.12364	2.056	0.039799	*
regocup4:gurb2	0.12356	0.16296	0.758	0.448330	
regocup2:gurb3	0.42370	0.08398	5.045	4.53e-07	***
regocup3:gurb3	0.40347	0.12996	3.105	0.001905	**
regocup4:gurb3	0.46526	0.14978	3.106	0.001894	**

Null deviance: 41364 on 29837 degrees of freedom
Residual deviance: 30013 on 29812 degrees of freedom
AIC: 30065

```
> anova(modf, modfinalstep, test="Chisq")
Analysis of Deviance Table

Model 1: desptot ~ rendtot + npagr + regocup + gurb + pais +
pais:rendtot +
  pais:npagr + pais:gurb + pais:regocup + rendtot:npagr +
npagr:regocup +
  gurb:regocup
Model 2: desptot ~ rendtot + npagr + regocup + gurb + pais +
pais:rendtot +
  pais:npagr + pais:gurb + pais:regocup + rendtot:npagr +
rendtot:regocup +
  rendtot:gurb + npagr:regocup + npagr:gurb + gurb:regocup
Resid. Df Resid. Dev Df Deviance P(>|Chi|)
1      29812      30013
2      29805      30012  7    1.0383    0.9942
```

```
> modf<-
glm(desptot~rendtot+npagr+regocup+gurb+pais+pais:rendtot+pais:npagr+pa
is:gurb+pais:regocup+rendtot:npagr+npagr:regocup+gurb:regocup,
family=binomial("logit"))
> summary(modf)
Call:
glm(formula = desptot ~ rendtot + npagr + regocup + gurb + pais +
  pais:rendtot + pais:npagr + pais:gurb + pais:regocup +
  rendtot:npagr +
  npagr:regocup + gurb:regocup, family = binomial("logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0495 -0.6213  0.5111  0.7354  2.3061

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.60980    0.05599 -28.752 < 2e-16 ***
rendtot1    2.54578    0.06125  41.564 < 2e-16 ***
npagr1      1.43753    0.06505  22.100 < 2e-16 ***
```

regocup2	0.57930	0.08467	6.842	7.83e-12	***
regocup3	-0.33105	0.08821	-3.753	0.000175	***
regocup4	-0.21513	0.13766	-1.563	0.118105	
gurb2	-0.40812	0.07313	-5.581	2.40e-08	***
gurb3	-0.79371	0.07560	-10.499	< 2e-16	***
pais1	0.06282	0.06605	0.951	0.341511	
rendtot1:pais1	-0.44342	0.06355	-6.977	3.01e-12	***
npagr1:pais1	0.18129	0.06466	2.804	0.005048	**
gurb2:pais1	0.18047	0.08172	2.208	0.027220	*
gurb3:pais1	0.19271	0.08320	2.316	0.020542	*
regocup2:pais1	-0.19878	0.08075	-2.461	0.013836	*
regocup3:pais1	-0.51100	0.09795	-5.217	1.82e-07	***
regocup4:pais1	-0.01509	0.13531	-0.111	0.911222	
rendtot1:npagr1	-0.54034	0.05986	-9.027	< 2e-16	***
npagr1:regocup2	-0.44296	0.07264	-6.098	1.07e-09	***
npagr1:regocup3	-0.29792	0.09343	-3.189	0.001430	**
npagr1:regocup4	-0.23405	0.12954	-1.807	0.070801	.
regocup2:gurb2	0.26346	0.08745	3.013	0.002591	**
regocup3:gurb2	0.25417	0.12364	2.056	0.039799	*
regocup4:gurb2	0.12356	0.16296	0.758	0.448330	
regocup2:gurb3	0.42370	0.08398	5.045	4.53e-07	***
regocup3:gurb3	0.40347	0.12996	3.105	0.001905	**
regocup4:gurb3	0.46526	0.14978	3.106	0.001894	**

Null deviance: 41364 on 29837 degrees of freedom
Residual deviance: 30013 on 29812 degrees of freedom
AIC: 30065

```

>Ponto de corte
> tp<-sum(desptot==1)
> tp
[1] 14921
> vp<-sum((modf$fitted>=0.53) & (desptot==1))
> vp
[1] 11182
> sensb<-vp/tp
> sensb*100
[1] 74.94136
> tn<-sum(desptot==0)
> tn
[1] 14917
> vn<-sum((modf$fitted<0.53) & (desptot==0))
> vn
[1] 11704
> esp<-vn/tn
> esp*100
[1] 78.46082

```